

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 29, 2005, 14:44:29 ; Search time 218 Seconds  
(without alignments)  
6214.849 Million cell updates/sec

Title: US-10-069-353A-7

Perfect score: 828  
Sequence: 1 gctgtccagctgcgcacc.....cgcggaagccgctgcgcga 828

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 8181359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/prodata/1/ina/5A COMB. seq: \*  
2: /cgn2\_6/prodata/1/ina/5B COMB. seq: \*  
3: /cgn2\_6/prodata/1/ina/6A COMB. seq: \*  
4: /cgn2\_6/prodata/1/ina/6B COMB. seq: \*  
5: /cgn2\_6/prodata/1/ina/PTUS COMB. seq: \*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	* Query Match	Length	DB	ID	Description
1	828	100.0	80161	3	US-09-036-987A-1	Sequence 1, Appl1
2	828	100.0	80161	3	US-09-370-700-1	Sequence 1, Appl1
3	828	100.0	80161	4	US-09-603-207-1	Sequence 1, Appl1
4	165.4	20.0	6085	3	US-09-029-603-4	Sequence 4, Appl1
5	165.4	20.0	80161	3	US-09-036-987A-1	Sequence 1, Appl1
6	165.4	20.0	80161	3	US-09-370-700-1	Sequence 1, Appl1
7	165.4	20.0	80161	4	US-09-603-207-1	Sequence 1, Appl1
8	133.2	16.1	852	4	US-09-266-965-36	Sequence 16, Appl1
9	133.2	16.1	53500	4	US-09-266-965-76	Sequence 76, Appl1
10	122.8	14.8	2122	3	US-09-029-603-1	Sequence 1, Appl1
11	83.2	10.0	828	4	US-09-266-965-35	Sequence 35, Appl1
12	71.2	8.6	4403765	3	US-09-103-840A-2	Sequence 2, Appl1
13	71.2	8.6	4411529	3	US-09-103-840A-1	Sequence 1, Appl1
14	63.2	7.6	1359	4	US-09-439-554-29	Sequence 29, Appl1
15	62.2	7.5	10263	4	US-09-902-540-1024	Sequence 1024, Ap
16	61	7.4	1131	4	US-09-252-991A-2322	Sequence 2322, Ap
17	61	7.4	1215	4	US-09-252-991A-2513	Sequence 2513, Ap
18	61	7.4	1215	4	US-09-252-991A-2595	Sequence 2595, Ap
19	60.4	7.3	991	4	US-09-902-540-3933	Sequence 3933, Ap
20	57.2	6.9	991	4	US-09-439-554-27	Sequence 27, Appl1
21	57.2	6.9	1152	4	US-09-614-221A-523	Sequence 523, App
22	54.8	6.6	507	4	US-09-439-554-13	Sequence 13, Appl1
23	51.4	6.2	669	4	US-09-902-540-9153	Sequence 9153, Ap
24	51.4	6.2	17245	4	US-09-902-540-1073	Sequence 1073, Ap
25	49.4	6.0	1212	4	US-09-252-991A-9135	Sequence 9135, Ap
26	49.4	6.0	1746	4	US-09-252-991A-8948	Sequence 8948, Ap
27	49.4	6.0	2304	4	US-09-252-991A-8731	Sequence 8731, Ap

28	48.8	5.9	507	4	US-09-902-540-7484	Sequence 7484, Ap
29	48.8	5.9	6599	4	US-09-902-540-724	Sequence 724, App
30	48.4	5.8	1428	4	US-09-902-540-8152	Sequence 8152, App
31	48.4	5.8	6269	4	US-09-902-540-836	Sequence 836, App
32	47.4	5.7	2370	1	US-08-104-072B-7	Sequence 7, Appl1
33	47.4	5.7	2370	1	US-08-351-413-8	Sequence 8, Appl1
34	47.4	5.7	2370	2	US-09-025-583-8	Sequence 8, Appl1
35	47.2	5.7	1893	4	US-09-252-991A-9055	Sequence 9055, Ap
36	47	5.7	4403765	3	US-09-103-840A-2	Sequence 2, Appl1
37	47	5.7	4411529	3	US-09-103-840A-1	Sequence 1, Appl1
38	45	5.4	783	4	US-09-252-991A-15973	Sequence 15973, A
39	45	5.4	933	4	US-09-252-991A-16068	Sequence 16068, A
40	44.8	5.4	510	4	US-09-252-991A-8834	Sequence 8834, Ap
41	44.2	5.3	957	3	US-09-382-906A-1	Sequence 1, Appl1
42	44.2	5.3	2985	3	US-08-724-814-18	Sequence 18, Appl1
43	44.2	5.3	4632	4	US-09-902-540-8368	Sequence 8368, Ap
44	44.2	5.3	15482	4	US-09-902-540-1067	Sequence 1067, Ap
45	44	5.3	597	4	US-09-252-991A-16384	Sequence 16384, A

## ALIGNMENTS

RESULT 1  
US-09-036-987A-1  
; Sequence 1, Application US/09036987A  
; Patent No. 6143526  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H.  
; APPLICANT: Broughton, Mary C.  
; APPLICANT: Crawford, Kathryn P.  
; APPLICANT: Maddur, Krishnamurthy  
; APPLICANT: Merlo, Donald J.  
; APPLICANT: Treadway, Patli J.  
; APPLICANT: Turner, Jan R.  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biochemical Genes For Spinosyn Insecticide  
; TITLE OF INVENTION: Production  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dow Agrosciences LLC Patent Department  
; STREET: 9330 Zionsville Road  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: USA  
; ZIP: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/036,987A  
; FILING DATE: 09-MAR-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stuart, Donald R.  
; REGISTRATION NUMBER: 28,479  
; REFERENCE/DOCKET NUMBER: 50,608  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (317)337-4816  
; TELEFAX: (317)337-4847  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 80161 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-09-036-987A-1  
Query Match 100.0%; Score 828; DB 3; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 1.7e-179;

```
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGTTCACAGTGGCCGACCAACATCGACAGAGTTGGGCAATGATGACCGTGTCAAG 60
Db 20168 GTGTTCACAGTGGCCGACCAACATCGACAGAGTTGGGCAATGATGACCGTGTCAAG 20227
QY 61 CCGTGTCTGAATCTCGGTTCGGGGGCGGCCCCCTGTGCGCATCCACCAGGCTACTGGAGAAC 120
Db 20228 CCGTGTCTGAATCTCGGTTCGGGGGCGGCCCCCTGTGCGCATCCACCAGGCTACTGGAGAAC 20287
QY 121 GACGGGGGGGCTTCCCTGCGAGAGAGGCGCGACCGGCTCAACGACCTTTTCCCGCAACGG 180
Db 20288 GACGGGGGGGCTTCCCTGCGAGAGAGGCGCGACCGGCTCAACGACCTTTTCCCGCAACGG 20347
QY 181 ACCGTGCTCATGAGCGCGCTGTGCACTGCTCATGTTGGGGTGGCGTACCGGACCAACGAGG 240
Db 20348 ACCGTGCTCATGAGCGCGCTGTGCACTGCTCATGTTGGGGTGGCGTACCGGACCAACGAGG 20407
QY 241 CTGCGCGTGCAGCGCGACCAACCGGATCCAGATCCAGGATCAACCGTCAACCGTCAAGGTTGCAA 300
Db 20408 CTGCGCGTGCAGCGCGACCAACCGGATCCAGATCCAGGATCAACCGTCAAGGTTGCAA 20467
QY 301 GTGGCCATGCGCGCGTATGGCGAGCGCGAACCGGATCTAAGCCACCGGGTGAATTTCTCG 360
Db 20468 GTGGCCATGCGCGCGTATGGCGAGCGCGAACCGGATCTAAGCCACCGGGTGAATTTCTCG 20527
QY 361 TCGCTGATGATGATGCTCCTGCGGTACCGCGAACATGCTTTGACGCGCGCTGGGCGCATG 420
Db 20528 TCGCTGATGATGATGCTCCTGCGGTACCGCGAACATGCTTTGACGCGCGCTGGGCGCATG 20587
QY 421 CAGTGGCTGTGAGATGTCCGAACCGGACCGTCCATCCGAGAAATCTTTCAGATCTC 480
Db 20588 CAGTGGCTGTGAGATGTCCGAACCGGACCGTCCATCCGAGAAATCTTTCAGATCTC 20647
QY 481 AAACCCGGGAGATCCTCGGCGTCAACCGAGTGTCAAAAGAGAGCGGCGCGCGGAGTGC 540
Db 20648 AAACCCGGGAGATCCTCGGCGTCAACCGAGTGTCAAAAGAGAGCGGCGCGGAGTGC 20707
QY 541 CCGGTGTCCGGGAGACAGTGGCGCAACCGGCTTTCGATGCTGCTGGAGCACTTCTG 600
Db 20708 CCGGTGTCCGGGAGACAGTGGCGCAACCGGCTTTCGATGCTGCTGGAGCACTTCTG 20767
QY 601 GAATGCTGTGCGTGCAGCGGGGTTTCCAGATCTTCGATTTGGAGAGACGTTGTCTCGAGAGAC 660
Db 20768 GAATGCTGTGCGTGCAGCGGGGTTTCCAGATCTTCGATTTGGAGAGACGTTGTCTCGAGAGAC 20827
QY 661 CCGTACTTCATGCGCGAGTTGCGCGAAGAGCTCGGTCGCGCACACAGACGCGGATCGCGAGC 720
Db 20828 CCGTACTTCATGCGCGAGTTGCGCGAAGAGCTCGGTCGCGCACACAGACGCGGATCGCGAGC 20887
QY 721 AGGTACGGGCGGCTGTGCGCGCTGGGCGCGCGGCTGTGCGATTATGAGAAATATATGCC 780
Db 20888 AGGTACGGGCGGCTGTGCGCGCTGGGCGCGCGGCTGTGCGATTATGAGAAATATATGCC 20947
QY 781 CACGACATGGGCTATGCGATTCTGAACGCGCGGAAAGCGGTTGGCTGA 828
Db 20948 CACGACATGGGCTATGCGATTCTGAACGCGCGGAAAGCGGTTGGCTGA 20995

RESULT 2
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patli J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 Div1
```

```
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Query Match          100.0%; Score 828; DB 3; Length 80161;
Best Local Similarity 100.0%; Pred. No. 1,7e-179;
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGTTCACAGTGGCCGACCAACATCGACAGAGTTGGGCAATGATGACCGTGTCAAG 60
Db 20168 GTGTTCACAGTGGCCGACCAACATCGACAGAGTTGGGCAATGATGACCGTGTCAAG 20227
QY 61 CCGTGTCTGAATCTCGGTTCGGGGGCGGCCCCCTGTGCGCATCCACCAGGCTACTGGAGAAC 120
Db 20228 CCGTGTCTGAATCTCGGTTCGGGGGCGGCCCCCTGTGCGCATCCACCAGGCTACTGGAGAAC 20287
QY 121 GACGGGGGGGCTTCCCTGCGAGAGAGGCGCGACCGGCTCAACGACCTTTTCCCGCAACGG 180
Db 20288 GACGGGGGGGCTTCCCTGCGAGAGAGGCGCGACCGGCTCAACGACCTTTTCCCGCAACGG 20347
QY 181 ACCGTGCTCATGAGCGCGCTGTGCACTGCTCATGTTGGGGTGGCGTACCGGACCAACGAGG 240
Db 20348 ACCGTGCTCATGAGCGCGCTGTGCACTGCTCATGTTGGGGTGGCGTACCGGACCAACGAGG 20407
QY 241 CTGCGCGTGCAGCGCGACCAACCGGATCCAGATCCAGGATCAACCGTCAACCGTCAAGGTTGCAA 300
Db 20408 CTGCGCGTGCAGCGCGACCAACCGGATCCAGATCCAGGATCAACCGTCAAGGTTGCAA 20467
QY 301 GTGGCCATGCGCGCGTATGGCGAGCGCGAACCGGATCTAAGCCACCGGGTGAATTTCTCG 360
Db 20468 GTGGCCATGCGCGCGTATGGCGAGCGCGAACCGGATCTAAGCCACCGGGTGAATTTCTCG 20527
QY 361 TCGCTGATGATGATGCTCCTGCGGTACCGCGAACATGCTTTGACGCGCGCTGGGCGCATG 420
Db 20528 TCGCTGATGATGATGCTCCTGCGGTACCGCGAACATGCTTTGACGCGCGCTGGGCGCATG 20587
QY 421 CAGTGGCTGTGAGATGTCCGAACCGGACCGTCCATCCGAGAAATCTTTCAGATCTC 480
Db 20588 CAGTGGCTGTGAGATGTCCGAACCGGACCGTCCATCCGAGAAATCTTTCAGATCTC 20647
QY 481 AAACCCGGGAGATCCTCGGCGTCAACCGAGTGTCAAAAGAGAGCGGCGCGGAGTGC 540
Db 20648 AAACCCGGGAGATCCTCGGCGTCAACCGAGTGTCAAAAGAGAGCGGCGCGGAGTGC 20707
QY 541 CCGGTGTCCGGGAGACAGTGGCGCAACCGGCTTTCGATGCTGCTGGAGCACTTCTG 600
Db 20708 CCGGTGTCCGGGAGACAGTGGCGCAACCGGCTTTCGATGCTGCTGGAGCACTTCTG 20767
QY 601 GAATGCTGTGCGTGCAGCGGGGTTTCCAGATCTTCGATTTGGAGAGACGTTGTCTCGAGAGAC 660
Db 20768 GAATGCTGTGCGTGCAGCGGGGTTTCCAGATCTTCGATTTGGAGAGACGTTGTCTCGAGAGAC 20827
QY 661 CCGTACTTCATGCGCGAGTTGCGCGAAGAGCTCGGTCGCGCACACAGACGCGGATCGCGAGC 720
Db 20828 CCGTACTTCATGCGCGAGTTGCGCGAAGAGCTCGGTCGCGCACACAGACGCGGATCGCGAGC 20887
QY 721 AGGTACGGGCGGCTGTGCGCGCTGGGCGCGCGGCTGTGCGATTATGAGAAATATATGCC 780
Db 20888 AGGTACGGGCGGCTGTGCGCGCTGGGCGCGCGGCTGTGCGATTATGAGAAATATATGCC 20947
QY 781 CACGACATGGGCTATGCGATTCTGAACGCGCGGAAAGCGGTTGGCTGA 828
Db 20948 CACGACATGGGCTATGCGATTCTGAACGCGCGGAAAGCGGTTGGCTGA 20995
```







SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 1  
 LENGTH: 80161  
 TYPE: DNA  
 ORGANISM: Saccharopolyspora spinosa  
 US-09-370-700-1

Query Match 20.0%; Score 165.4; DB 3; Length 80161;  
 Best Local Similarity 52.8%; Pred. No. 2.3e-28;  
 Matches 386; Conservative 0; Mismatches 336; Indels 9; Gaps 1;

15 CGACCAACATCGCAGAGTTGGGAGATGATGACCTGTCACCGCTTGTGAATC 74  
 13575 CACACAGGATTTCCGACCGCGGATCAGGTGAGTCACTTCGATGCTGGCGACCG 13516  
 75 GGTCCGGGGGCGCCCTCGCCATCCACACCGCTACCTGGAGAACGACGGGCGGCTTC 134  
 13515 GGTCCCGCTCGACACGAGTTACTGGGCGGGGTATCGGGAGGATGCGGTCCACACC 13456  
 135 CTGGCAGCAGCGCGCGGACCGGCTCACCGACTTGTGCGCGAACGACCGTGTCTGATGG 194  
 13455 GTGTGCGAGTGTGCGGACCACTGACCTGTTATCGACAGAGCGCGCTCGTCC 13396  
 195 CGGCGTTCAGCTGCTGATGTGGGGTGGCGGTACCGGACACAGCGCTGCGCGTCCGCG 254  
 13395 CGGAGCGCAGCTGTTGACCTGGGCTGGCGCATGGGCGACCCGTAGTCCGTGCGGATG 13336  
 255 CGACACCGGATCCAGATCAACCGGATCAACCGTACCGAGTCCAGTGGCCATCGCCG 314  
 13335 CGGACCGGGGTTTGAATCAACCGGATCAACCGTAAAGCCAGCATCTGCGCGCCGAC 13276  
 315 TGATTGCGCAGCGGAAAGCGGATTAAGCCACCGGGTGGATTTCTGCGGTGATGCCAT 374  
 13275 CAGGCTCGCCACGAGACCGGACCTGGCGGATGTTGATCTAGTCAACGCGCGC 13216  
 375 GTCCCTGCGGTACCGGACCAATGCTTTCAGCGCGCGCTGGCGCATGCTGTTGGA 434  
 13215 CAGCTGCGCTTACCGGACCGGATCTTTCAGCGCGCATGGGCGATGCTGCTGTC 13156  
 435 GATGTCGGAACCGGACCGGATCCATCCGGAATCTTTCAGATCTCAACCGGATGCGAT 494  
 13155 GATGTCGGAACCGGACCGGATCCATCCGGAATCTTTCAGATCTCAACCGGATGCGAT 13096  
 495 CTTGCGCGTCAACGAGGCTCTCAACGAGAACGGGGGCGGAGATGCGGTGTCGCGGA 554  
 13095 GTTCTGCTCGAGACATCACTCGGGTTCGATCTCCGGAAGATACCGCGCGGTTTG 13036  
 555 CAGGTGCGGACCGGCGCTTTCGATCTGCGGTGAGCACTTCTGGAATCGCTGCGTGC 614  
 13035 GACGGGACGACCGGCGCATCTT-----GACAGCTTCAACGCGCTGATCAGCA 12985  
 615 AGCGGGTTTCGAGATCTCGATTGGAGAGAGTGTGTCGAGACCGGATCTTCAATGCC 674  
 12984 AGCGGGTTTCGAGATCTCGATTGGAGAGAGTGTGTCGAGACCGGATCTTCAATGCC 12925  
 675 GCAGTTTCGCGAAGAGTCTGCTGCGACACGACCGGATGCGGACGATGCGGCGCGC 734  
 12924 CTGTGATGTCGAGAGATGCTGCTCGGAAACTGATGAGCTGCGCGCTGAGGCTGCGGC 12865  
 735 TGTGCGCGGCT 745  
 12864 TGTGCGCACCT 12854

RESULT 7  
 US-09-603-207-1/c  
 Sequence 1, Application US/09603207B  
 Patent No. 6521406  
 GENERAL INFORMATION:  
 APPLICANT: Baltz, Richard H  
 APPLICANT: Broughton, Mary C  
 APPLICANT: Crawford, Kathryn P  
 APPLICANT: Madduri, Krishnamurthy

APPLICANT: Treadway, Patli J  
 APPLICANT: Turner, Jan R  
 APPLICANT: Waldron, Clive  
 TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
 FILE REFERENCE: 50489 DIV1  
 CURRENT APPLICATION NUMBER: US/09/603,207B  
 EARLIER FILING DATE: 2000-06-23  
 EARLIER APPLICATION NUMBER: 09/370,700  
 NUMBER OF SEQ ID NOS: 39  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 1  
 LENGTH: 80161  
 TYPE: DNA  
 ORGANISM: Saccharopolyspora spinosa  
 US-09-603-207-1

Query Match 20.0%; Score 165.4; DB 4; Length 80161;  
 Best Local Similarity 52.8%; Pred. No. 2.3e-28;  
 Matches 386; Conservative 0; Mismatches 336; Indels 9; Gaps 1;

15 CGACCAACATCGCAGAGTTGGGAGATGATGACCTGTCACCGCTTGTGAATC 74  
 13575 CACACAGGATTTCCGACCGCGGATCAGGTGAGTCACTTCGATGCTGGCGACCG 13516  
 75 GGTCCGGGGGCGCCCTCGCCATCCACACCGCTACCTGGAGAACGACGGGCGGCTTC 134  
 13515 GGTCCCGCTCGACACGAGTTACTGGGCGGGGTATCGGGAGGATGCGGTCCACACC 13456  
 135 CTGGCAGCAGCGCGCGGACCGGCTCACCGACTTGTGCGCGAACGACCGTGTCTGATGG 194  
 13455 GTGTGCGAGTGTGCGGACCACTGACCTGTTATCGACAGAGCGCGCTCGTCC 13396  
 195 CGGCGTTCAGCTGCTGATGTGGGGTGGCGGTACCGGACACAGCGCTGCGCGTCCGCG 254  
 13395 CGGAGCGCAGCTGTTGACCTGGGCTGGCGCATGGGCGACCGGATGCTGCGGCGATG 13336  
 255 CGACACCGGATCCAGATCAACCGGATCAACCGTACCGAGTCCAGTGGCCATCGCCG 314  
 13335 CGGACCGGGGTTTGAATCAACCGGATCAACCGTAAAGCCAGCATCTGCGCGCCGAC 13276  
 315 TGATTGCGCAGCGGAAAGCGGATTAAGCCACCGGGTGGATTTCTGCGGTGATGCCAT 374  
 13275 CAGGCTCGCCACGAGACCGGATCCGAGTGGCGGATGTTGATGCTGATGTCAGCGCGC 13216  
 375 GTCCCTGCGGTACCGGACCAATGCTTTCAGCGCGCGCTGGCGCATGCTGTTGGA 434  
 13215 CAGCTGCGCTTACCGGACCGGATCTTTCAGCGCGCATGGGCGATGCTGCTGTC 13156  
 435 GATGTCGGAACCGGACCGGATCCATCCGGAATCTTTCAGATCTCAACCGGATGCGAT 494  
 13155 GATGTCGGAACCGGACCGGATCCATCCGGAATCTTTCAGATCTCAACCGGATGCGAT 13096  
 495 CTTGCGCGTCAACGAGGCTCTCAACGAGAACGGGGGCGGAGATGCGGTGTCGCGGA 554  
 13095 GTTCTGCTCGAGACATCACTCGGGTTCGATCTCCGGAAGATACCGCGCGGTTTG 13036  
 555 CAGGTGCGGACCGGCGCTTTCGATCTGCGGTGAGCACTTCTGGAATCGCTGCGTGC 614  
 13035 GACGGGACGACCGGCGCATCTT-----GACAGCTTCAACGCGCTGATCAGCA 12985  
 615 AGCGGGTTTCGAGATCTCGATTGGAGAGAGTGTGTCGAGACCGGATCTTCAATGCC 674  
 12984 AGCGGGTTTCGAGATCTCGATTGGAGAGAGTGTGTCGAGACCGGATCTTCAATGCC 12925  
 675 GCAGTTTCGCGAAGAGTCTGCTGCGACACGACCGGATGCGGACGATGCGGCGCGC 734  
 12924 CTGTGATGTCGAGAGATGCTGCTCGGAAACTGATGAGCTGCGCGCTGAGGCTGCGGC 12865  
 735 TGTGCGCGGCT 745  
 12864 TGTGCGCACCT 12854





QY	486	CGGTGGCATCCTCGGCGTCA	CCGAGTCTGT	515
Db	477	GGCGGCGCGGCTGACGTA	CGACTTCTGT	506

## RESULT 12

```

US-09-103-840A-2
: Sequence 2, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: TITLE OF INVENTION: TUBERCULOSIS
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 4403765
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: FEATURE:
: OTHER INFORMATION: CDC 1551
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
: US-09-103-840A-2

```

### RESULT 13

```

US-09-103-840A-1
/ Sequence 1, Application US/09103840A
/ Patent No. 6294328
/
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ TITLE OF INVENTION: TUBERCULOSIS
/
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A

```

```

: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ. ID NOS.: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ. ID NO. 1
: LENGTH: 441529
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: OTHER INFORMATION: H37Rv
US-03-103-840A-1

```

RESULT 14  
US-09-439-554-29

```

: Sequence 29 Application US/09439554
: Patent No. 6479733
: GENERAL INFORMATION:
: APPLICANT: Rafalski, Jan Antoni
: APPLICANT: Odell, Joan T.
: APPLICANT: Sakai, Hajime
: APPLICANT: Thorpe, Catherine J.
: APPLICANT: Kinney, Anthony J.
: APPLICANT: Famedu, Omolayo O.
: TITLE OF INVENTION: STEROL METABOLISM ENZYMES
: FILE REFERENCE: B8114 US NA
: CURRENT APPLICATION NUMBER: US/09/439,554
: CURRENT FILING DATE: 1999-11-12
: EARLIER APPLICATION NUMBER: 60/108,351
: EARLIER FILING DATE: 1998-No. 6479733ember-13
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: Microsoft Office 97
: SEQ ID NO 29
: LENGTH: 1369
: TYPE: DNA
: ORGANISM: Glycine max
: US-09-439-554-29

```

Position	Sequence	Position	Sequence
Db	446 CCCGGCAACAAATCTTGGACGTGGCTGGCGGGCCGACCATATGCGGGCCATTCGCG	505	
Qy	253 CGGCAACGCGATCCAGATCAACCGGCATCAACCGTCAGCCAGGTGCAAGTGCCATGCGC	312	
Db	506 GCCCATCTCCCGCGCGAAGGTGGGCAATCAACATCAACGATACAGGTCAATCGAGCA	565	
Qy	313 GCTGATTTGGCACCGCAACCGGACTAAGCCACCGGGTGAACCTTCTGTGTGTCATGCC	372	
Db	566 AGGATGCACAAACAAAGAGCTGGGTGGAATCTCTCTGTGCGAGGTCTGTGTGGGAATTTTC	625	
Qy	373 ATGTCCTGCGCCTAACCCCGGCAATGCTTTGCAAGCGCGCCTGGGACATGCTAGTCTGTTG	432	
Db	626 CTTAAGATGCTGTTTGTTCAGCAACAGCTTTCGACGGAAGGTACTCAATGAGGCCACGTGT	685	
Qy	433 GAGATGTCGAAACCGGACCGTGCCATCCGGGAAATCTTTCAGATCTCAAAACCGGTGGC	492	
Db	686 CACGCTCCCAAGCTGGAGAAGTGTACGCCGAATCTTCCGATCTTAAACCCGGGGCG	745	
Qy	493 ATCTCTGGGCTCACGAGGTCTGCAACGAGAAAGGGGCGGCGG	536	
Db	746 CTCTACGTTTCTAACAAGTGGGGTGAACAGACGCAATTAAGTACGGG	789	

```

RESULT 15
US-09-902-540-1024
; Sequence 1024, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1024
; LENGTH: 10263
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) ..(10263)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1024

```

Query Match	7.5%;	Score 62.2;	DB 4;	Length 10263;
Best Local Similarity	50.4%;	Pred. No. 5.8e-05;		
Matches 185;	Conservative	0;	Mismatches 173;	Indels 9;
				Gaps 1;

QY	141	GCAGGCGCCGACCGGGCTCACCAGACCTTGTGGCGGAAACGACCGGTCTGCATGGCGGGT	200
Db	1527	GCCACCTTTCGAGACGTTCTCTCAGGAGAGCCCTGTCTCCGGGCGGCGTGAAGCGCAAGGAC	1586
QY	201	TGCATGCTCGATGTGTGGGTGGGTGCGGTACCGGACAAACGACGGTGTGCGGTGCGGCGACAA	260
Db	1587	GCGTGTCTGTGACCGTGGCCACGGGGGCGGGGACCGTCCGCTGTGCTGCGCGCGCGACGG	1646
QY	261	CGCGATCCAGATACCGGGCATTCACCGTACGACGATGTGAATGGGCANTCGCGCGGTGATTG	320
Db	1647	CGC-----CGCGTGCACGACGTGATTCTCCCGGAGATGATTGCGGCGCTTGGC	1697
QY	321	CGACGCGAAACGCGGACTTAAGCCACCGGGTGAATCTTCGTGCGTCAATGCGCATGTCCCT	380
Db	1698	CGGCGGGAACCGCGGAGCGAGACTGACGTCGACATCTCTGGAGGGGTGACGGCATATGCGCT	1757
QY	381	GCGGTACCCGAGCATGCTTTTCGAGACCGGCTCGGGCCATTCGACATGCGCTGTTGGAGATGC	440
Db	1758	GCCCTTTCGAGGAGGAACGGGTTCGACGCGGCTTCTTCATATGTTTGGCGCTGATGTTCTTCC	1817

```

QY      441  CGAACCCGACCCGTGTCATCCGGGAAATCTTTCGAGTCTCAAAACCGGTGCATCTCGG  500
Db      1818  GGACCCGGGCAACGTGTCTTCAGAGCTTCACCCGGGTACTGAAGCCGGGCGGCCGT  1877

QY      501  CGTCACC  507
Db      1878  CGTGTCC  1884

Search completed: August 29, 2005, 17:11:53
Job time : 236 Secs

```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 29, 2005, 13:43:45 ; Search time 3886 Seconds  
(without alignments)  
10324.477 Million cell updates/sec

Title: US-10-069-353A-7

Perfect score: 828

Sequence: 1 ggtgtccagcggcggtcgccacc.....cgcggaagccggtcggtcga 828

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_bt: \*  
3: gb\_in: \*  
4: gb\_cm: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sta: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	828	100.0	828	6	AX089422 Sequence
2	828	100.0	45624	6	AX089419 Sequence
3	828	100.0	50000	6	AX089416 Sequence
4	828	100.0	80161	1	AY007564 Saccharop
5	828	100.0	80161	6	AR165018 Sequence
6	828	100.0	80161	6	BD137649 Biosynthe
7	828	100.0	80161	6	AR281866 Sequence
8	749.6	90.5	36538	6	AX600587 Sequence
9	170	20.5	846	6	AX598613 Sequence
10	170	20.5	52101	6	AX598593 Sequence
11	165.4	20.0	900	6	AX089434 Sequence
12	165.4	20.0	6085	6	AX0304 Sequence 4
13	165.4	20.0	6085	6	AR144763 Sequence
14	165.4	20.0	45624	6	AX089419 Sequence
15	165.4	20.0	50000	6	AX089416 Sequence
16	165.4	20.0	80161	1	AY007564 Saccharop
17	165.4	20.0	80161	6	AR165018 Sequence
18	165.4	20.0	80161	6	BD137649 Biosynthe
19	165.4	20.0	80161	6	AR281866 Sequence

20	159	19.2	36538	6	AX600587 Sequence
21	158.8	19.2	22999	1	AB088119 Streptom
22	142.2	17.2	210614	1	AB088224 Streptom
23	141.8	17.1	17512	1	AB071405 Lechevali
24	141.8	17.1	25681	1	SAE414559 Saccharo
25	141.8	17.1	26144	1	AB090952 Lechevali
26	141.8	17.1	28654	1	AF534707 Lechevali
27	140.8	17.0	846	1	AY396042 Streptom
28	140.6	17.0	109528	1	AF040570 Amycolato
29	137.6	16.6	5998	1	AB032524 Streptom
30	137.6	16.6	302675	1	AP005024 Streptom
31	133.2	16.1	852	6	BD269407 Streptom
32	133.2	16.1	852	6	AR266893 Sequence
33	133.2	16.1	53500	6	BD269447 Streptom
34	133.2	16.1	53500	6	AR266933 Sequence
35	133.2	16.1	63734	1	AF127374 Streptom
36	130	15.7	1410	6	E07846 DNA sequenc
37	122.8	14.8	2122	6	A60301 Sequence 1
38	122.8	14.8	2122	6	AR144762 Sequence
39	120.2	14.5	30000	6	AX50261 Sequence
40	120.2	14.5	103450	1	AF440781 Streptom
41	112.2	13.6	12381	6	BD217469 Polypeptid
42	112.2	13.6	12381	6	AX006889 Sequence
43	97.4	11.8	110000	8	CR382132_12 Continuation (13 o
44	95	11.5	12637	1	AF323753 Streptom
45	93.8	11.3	1585	8	AK060887 Oryza sat

#### ALIGNMENTS

RESULT 1	AX089422	828 bp	DNA	linear	PAT 21-MAR-2001
LOCUS	Sequence	7 from Patent WO0116303.			
DEFINITION	AX089422				
ACCESSION	AX089422.1	GI:13443683			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM	Saccharopolyspora spinosa				
	Saccharopolyspora spinosa				
	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
	Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.				
REFERENCE	1				
AUTHORS	Eberz, G., Moehrle, V., Froede, R., Velten, R. and Salas, J.A.				
TITLE	Nucleic acids which code for the enzyme activities of the spinosyn				
JOURNAL	bioSynthesis				
	Patent: WO 0116303-A 7 08-MAR-2001;				
FEATURES	LOCATION/Qualifiers				
source	1..828				
	/organism="Saccharopolyspora spinosa"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:60894"				
CDS	1..828				
	/note="unassigned protein product; ORF1; O-Methyltransferase"				
	/codon_start=1				
	/transl_table=11				
	/protein_id="CAC34874.1"				
	/db_xref="GI:13443684"				
	/translation="MPCGAPTSQGVQWYDVTPLNSVAGPCAIIHGWENDGRA				
	SWOQADRLTDLVAERTVLDVCGCGGAPALRAVANAIOITVSOOVA				
	IAADCARERGLSHRVDFSCVDAMSLPYPNAPDAWAMASLEMSPPRAIEIRLV				
	KRGIGITGVTRYKRRAGGMPYSGRWPTGIGICLAEOLESLRAAGPEILLMEVDS				
	KRYFMPQFAEBELAHQHGIAURYGPAVAGMAVCDIEKXAHMDGXIALTARKPVG"				
ORIGIN					
Query Match	100.0%; Score 828; DB 6; Length 828;				
Best Local Similarity	100.0%; Pred. No. 2.4e-116;				
Matches	828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1	GTTGTCAGGTCGCGCACCAATCGACGAGTTGGCAGATGTATGACCTGTCACG	60		
	1	GTTTCCAGGTGCGCCACCAATCGACGAGTTGGCAGATGTATGACCTGTCACG	60		
Db	1	GTTTCCAGGTGCGCCACCAATCGACGAGTTGGCAGATGTATGACCTGTCACG	60		

QY 61 CCGTTGTAACTCGGTCGGGGGCGGCCCTGCGCCATCCACGCGTACTGGAGAAC 120  
Db 61 CCGTTGTAACTCGGTCGGGGGCGGCCCTGCGCCATCCACGCGTACTGGAGAAC 120  
QY 121 GACGGGCGGGCTTCTCTGGACGAGGCGCGCCGACCCGGCTACCGACTTGTGCGCGAACG 180  
Db 121 GACGGGCGGGCTTCTCTGGACGAGGCGCGCCGACCCGGCTACCGACTTGTGCGCGAACG 180  
QY 181 ACCGTCGTGATGAGCGCGCTTGGACTGCTCCGATGTTGGGGTGTGGATACCGGACCAACGAG 240  
Db 181 ACCGTCGTGATGAGCGCGCTTGGACTGCTCCGATGTTGGGGTGTGGATACCGGACCAACGAG 240  
QY 241 CTGCGCGTCGCGCGACGACGAGATCCAGATCACCGGATCACCGTACCGAGGTCGAA 300  
Db 241 CTGCGCGTCGCGCGACGACGAGATCCAGATCACCGGATCACCGTACCGAGGTCGAA 300  
QY 301 GTGGCCATGCGCGCTGATTTGCGACGCGGAACTAAGCCACCGGGTGAATTCTTCG 360  
Db 301 GTGGCCATGCGCGCTGATTTGCGACGCGGAACTAAGCCACCGGGTGAATTCTTCG 360  
QY 361 TGGCGTCGATGCGATGCTTGCCTGCGATACCGGACCAATGCTTTGAGCGCGCGCTGGCGCATG 420  
Db 361 TGGCGTCGATGCGATGCTTGCCTGCGATACCGGACCAATGCTTTGAGCGCGCGCTGGCGCATG 420  
QY 421 CAGTCGCTGTTGAGAGATGTCGGAACCGGACCGTGGCCATCCGGGAAATCTTTCGAGTACTC 480  
Db 421 CAGTCGCTGTTGAGAGATGTCGGAACCGGACCGTGGCCATCCGGGAAATCTTTCGAGTACTC 480  
QY 481 AAACCCGGTGGCATCTCTCGGCGCTCACCGAGGTCGTCAAAGAGAAAGCGGGCGGGGATG 540  
Db 481 AAACCCGGTGGCATCTCTCGGCGCTCACCGAGGTCGTCAAAGAGAAAGCGGGCGGGGATG 540  
QY 541 CCGGTCGCGGGGACGAGTGGCGGACCGGGCTTCCGATGCTGCTGGTAGCAACTTCG 600  
Db 541 CCGGTCGCGGGGACGAGTGGCGGACCGGGCTTCCGATGCTGCTGGTAGCAACTTCG 600  
QY 601 GAATGCTGCTGACGAGCGGGGTTTCGAGATCTCTGATTTGGAGAGACGTCGTGAGAGACC 660  
Db 601 GAATGCTGCTGACGAGCGGGGTTTCGAGATCTCTGATTTGGAGAGACGTCGTGAGAGACC 660  
QY 661 CGGTACTTCATGCGCGGATTCGCGGAAAGCTGCTGCGCACGACGCGAGTCGCGAGC 720  
Db 661 CGGTACTTCATGCGCGGATTCGCGGAAAGCTGCTGCGCACGACGCGAGTCGCGAGC 720  
QY 721 AGGTACGGGCGCGCTGTCCCGCGCTGGGCGCGCGCTGTGGGATTTAGAAATATATGCC 780  
Db 721 AGGTACGGGCGCGCTGTCCCGCGCTGGGCGCGCGCTGTGGGATTTAGAAATATATGCC 780  
QY 781 CACGACATGGGCTATGCGATTCTGACGCGCGGAAAGCCGGTCGCTGA 828  
Db 781 CACGACATGGGCTATGCGATTCTGACGCGCGGAAAGCCGGTCGCTGA 828

RESULT 2  
AX089419/c 45624 bp DNA linear PAT 21-MAR-2001  
LOCUS AX089419 Sequence 4 from Patent WO0116303.  
DEFINITION AX089419  
ACCESSION AX089419  
VERSION AX089419.1 GI:13443680  
KEYWORDS  
SOURCE .  
ORGANISM Saccharopolyspora spinosa  
Saccharopolyspora spinosa  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardiales; Pseudonocardaceae; Saccharopolyspora.  
REFERENCE  
1 Eberz, G., Moehrl, V., Froede, R., Veltan, R. and Salas, J.A.  
Nucleic acids which code for the enzyme activities of the spinosyn  
biosynthesis  
JOURNAL Patent: WO 0116303-A 4 08-MAR-2001;  
AUTHORS BAYER AG (DE)  
FEATURES  
1. location/Qualifiers  
source

ORIGIN  
Query Match 100.0%; Score 828; DB 6; Length 45624;  
Best Local Similarity 100.0%; Pred. No. 9, se-117;  
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
/organism="Saccharopolyspora spinosa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:60894"

QY 1 GTGTTGCCAGGTGGGCGCAACAATCCGAGAGGTGGGCGAGTATGACTGTCACG 60  
Db 828 GTGTTGCCAGGTGGGCGCAACAATCCGAGAGGTGGGCGAGTATGACTGTCACG 769  
QY 61 CCGTTGTAACTCGGTCGGGGGCGGCCCTGCGCCATCCACGCGTACTGGAGAAC 120  
Db 768 CCGTTGTAACTCGGTCGGGGGCGGCCCTGCGCCATCCACGCGTACTGGAGAAC 709  
QY 121 GACGGGCGGGCTTCTCTGGACGAGGCGCGCCGACCCGGCTACCGGACTTGTGCGCGAACG 180  
Db 708 GACGGGCGGGCTTCTCTGGACGAGGCGCGCCGACCCGGCTACCGGACTTGTGCGCGAACG 649  
QY 181 ACCGTCGTGATGAGCGCGCTTGGACTGCTCCGATGTTGGGGTGTGGATACCGGACCAACGAG 240  
Db 648 ACCGTCGTGATGAGCGCGCTTGGACTGCTCCGATGTTGGGGTGTGGATACCGGACCAACGAG 589  
QY 241 CTGCGCGTCGCGCGACGACGAGATCCAGATCACCGGATCACCGTACCGAGGTCGAA 300  
Db 588 CTGCGCGTCGCGCGACGACGAGATCCAGATCACCGGATCACCGTACCGAGGTCGAA 529  
QY 301 GTGGCCATGCGCGCTGATTTGCGACGCGGAACTAAGCCACCGGGTGAATTCTTCG 360  
Db 528 GTGGCCATGCGCGCTGATTTGCGACGCGGAACTAAGCCACCGGGTGAATTCTTCG 469  
QY 361 TGGCGTCGATGCGATGCTTGCCTGCGATACCGGACCAATGCTTTGAGCGCGCGCTGGCGCATG 420  
Db 468 TGGCGTCGATGCGATGCTTGCCTGCGATACCGGACCAATGCTTTGAGCGCGCGCTGGCGCATG 409  
QY 421 CAGTCGCTGTTGAGAGATGTCGGAACCGGACCGTGGCCATCCGGGAAATCTTTCGAGTACTC 480  
Db 408 CAGTCGCTGTTGAGAGATGTCGGAACCGGACCGTGGCCATCCGGGAAATCTTTCGAGTACTC 349  
QY 481 AAACCCGGTGGCATCTCTCGGCGCTCACCGAGGTCGTCAAAGAGAAAGCGGGCGGGGATG 540  
Db 348 AAACCCGGTGGCATCTCTCGGCGCTCACCGAGGTCGTCAAAGAGAAAGCGGGCGGGGATG 289  
QY 541 CCGGTCGCGGGGACGAGTGGCGGACCGGGCTTCCGATGCTGCTGGTAGCAACTTCG 600  
Db 288 CCGGTCGCGGGGACGAGTGGCGGACCGGGCTTCCGATGCTGCTGGTAGCAACTTCG 229  
QY 601 GAATGCTGCTGACGAGCGGGGTTTCGAGATCTCTGATTTGGAGAGACGTCGTGAGAGACC 660  
Db 228 GAATGCTGCTGACGAGCGGGGTTTCGAGATCTCTGATTTGGAGAGACGTCGTGAGAGACC 169  
QY 661 CGGTACTTCATGCGCGGATTCGCGGAAAGCTGCTGCGCACGACGCGAGTCGCGAGC 720  
Db 168 CGGTACTTCATGCGCGGATTCGCGGAAAGCTGCTGCGCACGACGCGAGTCGCGAGC 109  
QY 721 AGGTACGGGCGCGCTGTCCCGCGCTGGGCGCGCGCTGTGGGATTTAGAAATATATGCC 780  
Db 108 AGGTACGGGCGCGCTGTCCCGCGCTGGGCGCGCGCTGTGGGATTTAGAAATATATGCC 49  
QY 781 CACGACATGGGCTATGCGATTCTGACGCGCGGAAAGCCGGTCGCTGA 828  
Db 48 CACGACATGGGCTATGCGATTCTGACGCGCGGAAAGCCGGTCGCTGA 1

RESULT 3  
AX089416 50000 bp DNA linear PAT 21-MAR-2001  
LOCUS AX089416 Sequence 1 from Patent WO0116303.  
DEFINITION AX089416  
ACCESSION AX089416  
VERSION AX089416.1 GI:13443677  
KEYWORDS



SOURCE	Saccharopolyspora spinosa									
ORGANISM	Saccharopolyspora spinosa									
REFERENCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardiales; Saccharopolyspora									
AUTHORS	1 Eberz, G., Moehle, V., Froede, R., Valten, R. and Salas, J.A.									
TITLE	Nucleic acids which code for the enzyme activities of the spinosyn biosynthesis									
JOURNAL	Patent: WO 0116303-A 1 08-MAR-2001;									
FEATURES	BAYER AG (DE)									
source	Location/Qualifiers									
	1..50000									
	/organism="Saccharopolyspora spinosa"									
	/mol_type="unassigned DNA"									
	/db_xref="taxon:60894"									
ORIGIN										
Query Match	100.0%; Score 828; DB 6; Length 50000;									
Best Local Similarity	100.0%; Pred. No. 9, 3e-117;									
Matches	828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	GTGTTGCCAGGTGGCGCAGCAACATCGCAGCAGGTTGGGAGATGATGATACCTCGTCAAG	60							
Db	44797	GTGTTGGCAGGTGGCGCAGCAACATCGCAGCAGGTTGGGAGATGATGATACCTCGTCAAG	44856							
QY	61	CCGTTGCTGAACCTGGTGTGGGGGGGGCCCTGGGGCCATCCACACAGGCTACTAGGGAGAAC	120							
Db	44857	CCGTTGCTGAACCTGGTGTGGGGGGGGCCCTGGGGCCATCCACACAGGCTACTAGGGAGAAC	44916							
QY	121	GACGGGGGGGCTTCTTGCGCAGCAGGCCCGCAGACCGGCTCACCGACCTTGTCGCGAACG	180							
Db	44917	GACGGGGGGGCTTCTTGCGCAGCAGGCCCGCAGACCGGCTCACCGACCTTGTCGCGAACG	44976							
QY	181	ACCGTGTCCATGCGCGCGCTTTCATCTCTGATGTGGGGTGGCGGTACCGGACCAACAGCG	240							
Db	44977	ACCGTGTCTGATGCGCGCGCTTTCATCTCTGATGTGGGGTGGCGGTACCGGACCAACAGCG	45036							
QY	241	CTGGCGGTGGCGCGCAGCAACAGCGCATCCAGATCACCGGCTACCGGTACGCGAGTGCA	300							
Db	45037	CTGGCGGTGGCGCGCAGCAACAGCGCATCCAGATCACCGGCTACCGGTACGCGAGTGCA	45096							
QY	301	GTGGCCATCGCGCGTATTGGCGACGGGAAACGGCGACTAAGCACCAGGTGACCTTCTCG	360							
Db	45097	GTGGCCATCGCGCGTATTGGCGACGGGAAACGGCGACTAAGCACCAGGTGACCTTCTCG	45156							
QY	361	TGCGTGCATGCGCATGTCCCTCCGTACCCGAGCAATCTTTCAACGCCCGCTGGGGCAG	420							
Db	45157	TGCGTGCATGCGCATGTCCCTCCGTACCCGAGCAATCTTTCAACGCCCGCTGGGGCAG	45216							
QY	421	CAGTGTGCTTGGAGATGTCCGAAACCGGACCGGTCCCATCCGGGAAATCTTTCAGATCTC	480							
Db	45217	CAGTGTGCTTGGAGATGTCCGAAACCGGACCGGTCCCATCCGGGAAATCTTTCAGATCTC	45276							
QY	481	AAACCCGGTGGCATCTCTCGCGCTCACCGAGTCTCAACAGAAAGCGGGCGGGGATG	540							
Db	45277	AAACCCGGTGGCATCTCTCGCGCTCACCGAGTCTCAACAGAAAGCGGGCGGGGATG	45336							
QY	541	CCGGTGTTCGGGGAGACAGGTGGCCGACCGGGCTTGGATCTGCTCGGTAGGCAATTTTG	600							
Db	45337	CCGGTGTTCGGGGAGACAGGTGGCCGACCGGGCTTGGATCTGCTCGGTAGGCAATTTTG	45396							
QY	601	GAATGTGCTGTGACGCGGGGTTGAGATCTCTGATTGGAGAGACGTGTGTGAGAGACC	660							
Db	45397	GAATGTGCTGTGACGCGGGGTTGAGATCTCTGATTGGAGAGACGTGTGTGAGAGACC	45456							
QY	661	CGGTACTTCAATGCGCAGTGTTCGCGAAGAGCTGCTCGGACCAAGCAGCGGATCGGGAC	720							
Db	45457	CGGTACTTCAATGCGCAGTGTTCGCGAAGAGCTGCTCGGACCAAGCAGCGGATCGGGAC	45516							
QY	721	AGGTACGGGGCGGGCTGTGCGCGGCGCGGGCGCGGGTCTGCGCAATTATGAGAAATATGCC	780							
Db	45517	AGGTACGGGGCGGGCTGTGCGCGGCGCGGGCGCGGGTCTGCGCAATTATGAGAAATATGCC	45576							

Db	Accession	Definition	Source
Oy	781	CACGACATGGCGCTATGCATTTCGACGCCGCGAGCCGGTTCGGCTTA	828
Db	45577	CACGACATGGCGCTATGCATTTCGACGCCGCGAGCCGGTTCGGCTTA	45624
RESULT 4 LOCUS AY007564	AY007564	80161 bp DNA linear BCT 16-MAY-2001	
DEFINITION	Saccharopolyspora spinosa probable lysR-type transcriptional regulator, probable keto acyl reductase, spinosad biosynthetic gene cluster, complete sequence, and probable exodeoxyribonuclease V genes, complete cds; and unknown gene.		
KEYWORDS	AY007564		
VERSION	AY007564.1	GI:13162633	
ORGANISM	Saccharopolyspora spinosa		
REFERENCE	Waldron, C., Matsushima, P., Rostek, P.R. Jr., Broughton, M.C., Turner, J., Madduri, K., Crawford, K.P., Merlo, D.J. and Baltz, R.H.		
AUTHORS	Cloning and analysis of the spinosad biosynthetic gene cluster of Saccharopolyspora spinosa		
TITLE	Chem. Biol. 8 (5), 487-499 (2001)		
JOURNAL	21257765		
MEDLINE	11358695		
PUBMED	2 (bases 1 to 80161)		
REFERENCE	Waldron, C., Matsushima, P., Rostek, P.R. Jr., Broughton, M.C., Turner, J., Madduri, K., Crawford, K.P., Merlo, D.J. and Baltz, R.H.		
AUTHORS	Direct Submission		
TITLE	Submitted (18-AUG-2000) Dow Agrosciences, 9330 Zionville Rd., Indianapolis, IN 46268, USA		
JOURNAL	Location/Qualifiers		
FEATURES	1..80161		
source	/organism="Saccharopolyspora spinosa"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:60894"		
	complement(1135..1971)		
	/note="ORF L16"		
	/codon_start=1		
	/transl_table=1		
	/product="probable lysR-type transcriptional regulator"		
	/protein_id="AAG33282.1"		
	/db_xref="gi:13162634"		
	/translation="METRELYRFVAVAEELIFGRAQRLGIAOPLSRITAAQLBQRLLGLIORTSRKSLTEAGAMILLTGRAILIGALAAERRTQATSPSVILAAGASGVALLAKIDAVAAEPGAVALVDLICEPOKTIHGADVALHOPFDTABLDIEINTNLEGEVTLTPSHPLASEPHRMADVSSLPDLFLAWMGPGCVYDGGVEVRNTOLFE		
	TEOOVALIPHSHPPLASEPHRMADVSSLPDLFLAWMGPGCVYDGGVEVRNTOLFE		
	OMIALGRITVMPESSRVNLLEGAAVPVLDAPVTTVIAMPHPSRBALAGLVAVAT		
	L16"		
	2024..2791		
	/note="ORF L15"		
	/codon_start=1		
	/transl_table=1		
	/product="probable keto acyl reductase"		
	/protein_id="AAG33281.1"		
	/db_xref="gi:13162633"		
	/translation="MWLVPRRRFRATLSMSBOTIALVTGANKGIGEYIAAGLAGMWSAGIARPDHOREGDVAVKIRADGVDAFVSILDVDTDSVAALAALEBRAGLDVLNNHVGISARDPHEESTPTPASLRVAVETNYIGVRYTNMMLPILRSEBRPRIYNQSSHVASLHTLTDPGVLDGIGSIGSAVSRSKTFLNLIITIOYAELSSTNIKINNACGYATLDNGFHGTSTPADGARIALRIALTLPDDGPPTGMFFDAGNPWM"		
	complement(3416..4165)		
	/gene="spns"		
	complement(3416..4165)		
	/gene="spns"		
	complement(3416..4165)		
	/gene="spns"		
	/note="probably involved in forosamine biosynthesis"		
	/codon_start=1		
	/evidence-not experimental		
	/transl_table=1		
	/product="probable N-dimethyltransferase"		
	/protein_id="AAG32280.1"		

gene /db\_xref="GI:13162652"  
/translation="MSRVSDFEATSTSVSPDHADIYDAIHSAGRDMAAGEVVOU  
VTRLEPAOSLIDVACGTGATERRAEKAYAGLEISDPAKEIAIRRPEVPIHGD  
IRDFLGEPEFDVITGLCTFAAMRTVDLDRTRMAMARHLAGVAVIEPMPKPI  
DGFVAHAHGERVIRSLSHSVLEGRTSRMVRTVAPEPTGIRPTEFEILSTED  
ETVALDEDAGIABYLPGAPNGRGLFVGIRN"  
/complement (4168..5325)  
/gene="spn"  
/complement (4168..5325)  
/gene="spn"  
/complement (4168..5325)  
/note="probably involved in forosamine biosynthesis"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="probable aminotransferase"  
/protein\_id="AAG23279.1"  
/db\_xref="GI:13162651"  
/translation="WIMLHQPLTGBEDDAIEVNASWITGCPRTTPEAFNHLG  
VDEQVFLNLSGTALFLTVQVLDLGPDDVLPISVPAANALASSGARPFCDVD  
PRLNPTLDDVRAITTPATKAVILAHYSGSPGEVTAIDFCREKGLMIEDSCAVAS  
SVHGACGFGDLATMSFDAMKILVYGGMFYADPELAHRARLAHYGLEOMGFD  
SAKSNRMWDIREDIGORLIGDMTALGSVOLKRLPEFIRRRRIATOVDRILSDV  
BGVLLPRLPDGVSSHVYVYQVLAPELRDOVAQOMLERGITSTRYPPLKVPITYRA  
DCKLSAEDACKRTLLPLHPSLDDAEVTVADPEOKAVEHHISQSPURK"  
/complement (5363..6751)  
/gene="spn"  
/complement (5363..6751)  
/note="probably involved in forosamine biosynthesis"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="probable NDP-hexose-3,4-dehydratase"  
/protein\_id="AAG23278.1"  
/db\_xref="GI:13162650"  
/translation="WQSRKTRALGKGRARVNSCDTTCATTEMVDAADRIIASVRY  
HREQSPFVAGSTPIRPSGAVLDEDDVALVEALEIRIAAGNARFESFAFVG  
LRKALVNSGSANLIALSLSPKLGARLPGEVITTAVGFTTINPVOGLVP  
VFDVDELGVNATPDIRKAAYETRAIMLANTLGNPPAADEIAIAKEHELFVEDN  
CDVSGSTRGLTGFEGDLTVVSPFAHHITSGEGCVLTGSLRLARIIESLRDGRD  
CMCEGVNCTCKRFPDYLGTLPQYDHKTFESHVGNLKTDDLOALALQLSKISA  
FGSARRRWRRLRELSGLPGILLPVATPHSPSPFGAITTSADAGTRALVNFLE  
SNIGTRLFGSNITRHPAFBOVRYRIADALINSDI VTRTFWGVVPGITDQMDIVY  
VESIAFVAKSS"  
/complement (7083..8450)  
/gene="spn"  
/complement (7083..8450)  
/note="involved in forosamine addition"  
/codon\_start=1  
/transl\_table=1  
/product="probable NDP-forosamyItransferase"  
/protein\_id="AAG23277.1"  
/db\_xref="GI:13162649"  
/translation="WIDMLPGCSIAIEBNKVLFTPLPASHFENLVPLAMALPAAG  
HEVRVAICPNWVMTGAGLTAVPVGDLDLISLAANKELVSGSVSDEGRHELF  
DELSINSGRDITAVEQLHLVDRLDDLMGAEKMQPDLVWDMAMVSCGVVAALG  
ARHVMLVALDVSGLRSFLEYSRKEPEORVDPDGLWGLAKLAFATPDEEVLG  
QATIDTPSMWRLPYDLDIYMRFPVNGPAVLTASDEVRMGEIPBNVYHEVPL  
RVTBQYEQSDOEQAMVERLLRGARLIDVEYATISDEVRMGEIPBNVYHEVPL  
NELRESGVIIHSGTTOETRTVNGVPLILPGTFWESRRAEILADRGAGLVADPA  
TFTEDVDRGQLARLIDEPSFANALIRREIESSPHDIVRLEKLVABERNRTGQ  
SDGHP"  
/complement (8967..10427)  
/gene="spn"  
/complement (8967..10427)  
/note="involved in forosamine biosynthesis"  
/codon\_start=1  
/transl\_table=1  
/product="probable NDP-hexose-2,3-dehydratase"  
/protein\_id="AAG23276.1"  
/db\_xref="GI:13162648"

gene /translation="WSSSVENASAAAPLGSNNTRRFVDSALSAQNGMIPTEFHCWL  
ADRGKNSFETNRILPDRLSKMKDASTENLVHADGFPFTEGQVETNRYGAASMQ  
PILOAEKGIIVLEKIDVGLHGMKMEKNGVNVQLSPTVOATSNVYAGASMG  
VPPVVDYTGGRVRVLDVTCOSQSGFVYKRRNMVNEQVEVPVLPDSCMLTIGQ  
VLALLRQDNTYNNTRTYLSTCIPPHDSATGTELAASEPFPQAVARSLSHIDSSSIS  
EAVMFESAKARYRLRATRVPLSVKMYRTDTSIAHODGYFAVIAVSATREVA  
SMTOPMIEPREQGBIALVKRIGVLLHVARVEAGYKMAEIAIAPVQCSVANYST  
PSNDMPPELDDVLTADPEVYVESILSESGRFYQAOVRRIIEVHEDFAARPPSDFR  
MWTIGQGEILSRSTHFLNIQARSLVASLHSLMALGR"  
/complement (10436..11434)  
/gene="spn"  
/complement (10436..11434)  
/note="involved in forosamine biosynthesis"  
/codon\_start=1  
/transl\_table=1  
/product="probable NDP-hexose-3-ketoreductase"  
/protein\_id="AAG23275.1"  
/db\_xref="GI:13162647"  
/translation="MRKPVRIQVGCASFARRMIPAMCDVAETEVAVASRDPKAE  
RFAARFEGEAVLQORLEERDIDAVVPLPQGNHAEIGKALPADKHVLAKEPLTTT  
ASDPAFLVGLARRNLLIRENYFLHHGRHVVRDILQSGEISGLAREFTAVGIPLP  
DTRIRYRELGGGLLDIGVYPARAHHPLIGPLTVYGASSHHEQSGVDSSVULO  
SEGGTVHLGTFVHHRSAVELMGSRGRLVDRAPFPPAWQVATIRKGVDELS  
LPAEDQVRKAVTAARDIRAGTGVDDPAVDAGSESMIQOALVLEAIGQARRCGST"  
/complement (11530..12492)  
/gene="spn"  
/complement (11530..12492)  
/note="involved in spinosyn aglycone biosynthesis"  
/codon\_start=1  
/transl\_table=1  
/product="SpnM"  
/protein\_id="AAG23274.1"  
/db\_xref="GI:13162646"  
/translation="MPNAGSTVLPVPIPMPREPRITTRAVGTHGISOVAPEYLK  
TGTEPETELVALADRGMAVITDYEGLPFGHTVYVGRAGQAMLDAAQAARLFG  
SGLTITDCEPVGIMGAQGOASAPFGEHLPTVAPBLRPAQAAGVPIDLDDIIRNDG  
VFTRPVLAAGVGAAYPDLPEFDELTEAGRTAVDQVRELCAPALVTRFLGRLESDL  
DTSGLPPORBARLARSVAGRNCPVPLTVHSTODEIVPFAFGRLRDSYAAAT  
PVRRHPLSGHAPPAALASSRNVVSWMDHSEBSAISGRDAR"  
/complement (12696..13547)  
/gene="spn"  
/complement (12696..13547)  
/note="involved in spinosyn aglycone biosynthesis;  
possibly involved in polyketide bridging"  
/codon\_start=1  
/transl\_table=1  
/product="methyltransferase-1-like protein"  
/protein\_id="AAG23273.1"  
/db\_xref="GI:13162645"  
/translation="MESIFDALHAGRELHNGVYAGRYDAGATPWSADADQJLDLFI  
DKALRFGAHLFDLGCNGQGVYVRAACSGVTVGTIVNAOHLAATRLANETLAS  
LEFDLVGAQDLPYDGFQQAAMQSVQVLDQAAALREHRIILEPGRFVLGDIITR  
VLEPEERAAVMTGTAHTLNSFTLVEAGEEILIEVLDLAQTRCWSWYVDELLRLK  
DELAGVPAVAVGTQQRVLDIAAKHGPPAQLIAAAYERKHPDYARNESMGFMLL  
QARKQGS"  
/complement (13592..14785)  
/gene="spn"  
/complement (13592..14785)

Query Match 100.0%; Score 828; DB 1; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 8,3e-117;  
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGTTGCCAGGTGCGCACCAATCGACAGGTTG3GGCAGATGTATGACCTGTCACG 60  
|||  
Db 20168 GTGTTGCCAGGTGCGCACCAATCGACAGGTTG3GGCAGATGTATGACCTGTCACG 20227  
|||  
Qy 61 CCGTTGCTGAACCTGGTCGGGGGGGGCCCTGGCCCAATCCACCAACGGCTACTGGGAAC 120  
|||  
Db 20228 CCGTTGCTGAACCTGGTCGGGGGGGGCCCTGGCCCAATCCACCAACGGCTACTGGGAAC 20287  
|||

QY 121 GACGGGCGGAGCTTCTGAGACAGAGCCGCGACCGGCTGACCGACTTGTGCGCGAAGCG 180  
 DB 20288 GACGGGCGGAGCTTCTGAGACAGAGCCGCGACCGGCTGACCGACTTGTGCGCGAAGCG 20347  
 QY 181 ACCGCTGCTGATGAGCGGCGGCTGAGCTGATGAGGAGGAGGAGTACCGGACCAACAGCG 240  
 DB 20348 ACCGCTGCTGATGAGCGGCGGCTGAGCTGATGAGGAGGAGGAGTACCGGACCAACAGCG 20407  
 QY 241 CTGCGGCTGCGCGCGACCAACCGGATCCAGATCACCAGGATCACTGTCAGGACAGTGCAG 300  
 DB 20408 CTGCGGCTGCGCGCGACCAACCGGATCCAGATCACCAGGATCACTGTCAGGACAGTGCAG 20467  
 QY 301 GTGGCCATGCGCGCTGATGAGCGGACCGGACCTGAGTAAAGCAACCGGAGTGAATCTTCG 360  
 DB 20468 GTGGCCATGCGCGCTGATGAGCGGACCGGACCTGAGTAAAGCAACCGGAGTGAATCTTCG 20527  
 QY 361 TGGGCTGATGAGCAATGCTCCGCGCTGACCGGACCAATGCTTTCGAGCGCGCTGAGGACATG 420  
 DB 20528 TGGGCTGATGAGCAATGCTCCGCGCTGACCGGACCAATGCTTTCGAGCGCGCTGAGGACATG 20587  
 QY 421 CAGTCCCTGTTGAGATGATCCGACCAACCGGACCGGATCCGAGGAAATCCTTCGAGTACTC 480  
 DB 20588 CAGTCCCTGTTGAGATGATCCGACCAACCGGACCGGATCCGAGGAAATCCTTCGAGTACTC 20647  
 QY 481 AAACCCGGTGGCATCTCTCGGCGTCAACCGGATGCTCAACGAGAACCGGCGCGGAGATG 540  
 DB 20648 AAACCCGGTGGCATCTCTCGGCGTCAACCGGATGCTCAACGAGAACCGGCGCGGAGATG 20707  
 QY 541 CCGGTCCTCGGAGGACAGGTCGCGGACCGGCTTCGATGCTGCTGAGTGAATCTTCG 600  
 DB 20708 CCGGTCCTCGGAGGACAGGTCGCGGACCGGCTTCGATGCTGCTGAGTGAATCTTCG 20767  
 QY 601 GAATCCCTGCTGAGCGGAGGTCGAGATCTCGATTTGGAGGACGTCGTGAGAGACC 660  
 DB 20768 GAATCCCTGCTGAGCGGAGGTCGAGATCTCGATTTGGAGGACGTCGTGAGAGACC 20827  
 QY 661 CCGTACTTCATGCGCGAGTTCGCGCAAGAGCTGCTGCGCACCAAGACCGGATCCGCGAC 720  
 DB 20828 CCGTACTTCATGCGCGAGTTCGCGCAAGAGCTGCTGCGCACCAAGACCGGATCCGCGAC 20887  
 QY 721 AGGTACGGGCGGCTGCTGCGGCGCTGGGCGCGCGGCTGCGGATTAAGAAATATGCC 780  
 DB 20888 AGGTACGGGCGGCTGCTGCGGCGCTGGGCGCGCGGCTGCGGATTAAGAAATATGCC 20947  
 QY 781 CACGACATGGGCTATGCGATTTGACGCGCGGAAAGCCGCTCGCTGA 828  
 DB 20948 CACGACATGGGCTATGCGATTTGACGCGCGGAAAGCCGCTCGCTGA 20995

RESULT 5  
 ARI65018 80161 bp DNA linear PAT 17-OCT-2001  
 LOCUS Sequence 1 from patent US 6274350.  
 DEFINITION ARI65018  
 ACCESSION ARI65018  
 VERSION ARI65018.1 GI:16238344  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 Unclassified.

REFERENCE 1 (bases 1 to 80161)  
 AUTHORS Baitz,R.H., Crawford,K.P., Broughton,M,Christine., Madduri,K.,  
 Merlo,D.J., Turner,J.R., Treadway,P.J., and Waldron,C.  
 TITLE Biosynthetic genes for spinosyn insecticide production  
 JOURNAL Patent: US 6274350-A 14-AUG-2001;  
 FEATURES  
 location/Qualifiers  
 1..80161  
 /organism="unknown"  
 /mol\_type="unassigned DNA"

ORIGIN  
 Query Match 100.0%; Score 828; DB 6; Length 80161;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-117;

Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTTGCCAGGTGGGCGACCAATCGAGAGGTTGGGAGATGATGACCTGCTGACG 60  
 DB 20168 GTGTTGCCAGGTGGGCGACCAATCGAGAGGTTGGGAGATGATGACCTGCTGACG 20227  
 QY 61 CCGTTGCTGAATCGGTCGCGGCGGCGGCGCTGCGCATCAACGAGCTACTGAGAGAAC 120  
 DB 20228 CCGTTGCTGAATCGGTCGCGGCGGCGGCGCTGCGCATCAACGAGCTACTGAGAGAAC 20287  
 QY 121 GACGGGCGGAGCTTCTGAGACAGAGCCGCGACCGGCTGACCGACTTGTGCGCGAAGCG 180  
 DB 20288 GACGGGCGGAGCTTCTGAGACAGAGCCGCGACCGGCTGACCGACTTGTGCGCGAAGCG 20347  
 QY 181 ACCGCTGCTGATGAGCGGCGGCTGAGCTGATGAGGAGGAGGAGTACCGGACCAACAGCG 240  
 DB 20348 ACCGCTGCTGATGAGCGGCGGCTGAGCTGATGAGGAGGAGGAGTACCGGACCAACAGCG 20407  
 QY 241 CTGCGGCTGCGCGCGACCAACCGGATCCAGATCACCAGGATCACTGTCAGGACAGTGCAG 300  
 DB 20408 CTGCGGCTGCGCGCGACCAACCGGATCCAGATCACCAGGATCACTGTCAGGACAGTGCAG 20467  
 QY 301 GTGGCCATGCGCGCTGATGAGCGGACCGGACCTGAGTAAAGCAACCGGAGTGAATCTTCG 360  
 DB 20468 GTGGCCATGCGCGCTGATGAGCGGACCGGACCTGAGTAAAGCAACCGGAGTGAATCTTCG 20527  
 QY 361 TGGGCTGATGAGCAATGCTCCGCGCTGACCGGACCAATGCTTTCGAGCGCGCTGAGGACATG 420  
 DB 20528 TGGGCTGATGAGCAATGCTCCGCGCTGACCGGACCAATGCTTTCGAGCGCGCTGAGGACATG 20587  
 QY 421 CAGTCCCTGTTGAGATGATCCGACCAACCGGACCGGATCCGAGGAAATCCTTCGAGTACTC 480  
 DB 20588 CAGTCCCTGTTGAGATGATCCGACCAACCGGACCGGATCCGAGGAAATCCTTCGAGTACTC 20647  
 QY 481 AAACCCGGTGGCATCTCTCGGCGTCAACCGGATGCTCAACGAGAACCGGCGCGGAGATG 540  
 DB 20648 AAACCCGGTGGCATCTCTCGGCGTCAACCGGATGCTCAACGAGAACCGGCGCGGAGATG 20707  
 QY 541 CCGGTCCTCGGAGGACAGGTCGCGGACCGGCTTCGATGCTGCTGAGTGAATCTTCG 600  
 DB 20708 CCGGTCCTCGGAGGACAGGTCGCGGACCGGCTTCGATGCTGCTGAGTGAATCTTCG 20767  
 QY 601 GAATCCCTGCTGAGCGGAGGTCGAGATCTCGATTTGGAGGACGTCGTGAGAGACC 660  
 DB 20768 GAATCCCTGCTGAGCGGAGGTCGAGATCTCGATTTGGAGGACGTCGTGAGAGACC 20827  
 QY 661 CCGTACTTCATGCGCGAGTTCGCGCAAGAGCTGCTGCGCACCAAGACCGGATCCGCGAC 720  
 DB 20828 CCGTACTTCATGCGCGAGTTCGCGCAAGAGCTGCTGCGCACCAAGACCGGATCCGCGAC 20887  
 QY 721 AGGTACGGGCGGCTGCTGCGGCGCTGGGCGCGCGGCTGCGGATTAAGAAATATGCC 780  
 DB 20888 AGGTACGGGCGGCTGCTGCGGCGCTGGGCGCGCGGCTGCGGATTAAGAAATATGCC 20947  
 QY 781 CACGACATGGGCTATGCGATTTGACGCGCGGAAAGCCGCTCGCTGA 828  
 DB 20948 CACGACATGGGCTATGCGATTTGACGCGCGGAAAGCCGCTCGCTGA 20995

RESULT 6  
 BD137649 80161 bp DNA linear PAT 18-SEP-2002  
 LOCUS BD137649  
 DEFINITION Biosynthetic genes for spinosyn insecticide production.  
 ACCESSION BD137649  
 VERSION BD137649.1 GI:23232594  
 KEYWORDS JP 2002505881-A/1.  
 SOURCE Saccharopolyspora spinosa  
 ORGANISM Saccharopolyspora spinosa  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Pseudonocardiales; Pseudonocardaceae; Saccharopolyspora.  
 REFERENCE 1 (bases 1 to 80161)  
 AUTHORS Baitz,R.H., Broughton,M.C., Crawford,K.P., Madduri,K., Merlo,D.J.,  
 Treadway,P.J., Turner,J.R. and Waldron,C.

TITLE Biosynthetic genes for spinosyn insecticide production  
JOURNAL Patent: JP 2002505891-A 1 26-FEB-2002;  
DOW AGROSCIENCES LLC  
COMMENT OS Saccharopolyspora spinosa  
PN JP 2002505891-A/1  
PD 26-FEB-2002  
PF 16-FEB-1999 JP 2000535754  
PR 09-MAR-1998 US 09/036987  
PI RICHARD H BALTZ, M CHRISTINE BROUGHTON, KATHRYN P CRAWFORD, PI  
KRISHNAMURTHY MADDURI, DONALD J MERLO, PATTI J TREADWAY, JAN R PI  
TURNER  
PI CLIVE WALDRON  
PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12P19/62, C12Q1/  
PC 68//  
PC C07K14/195, C12N15/00, C12N5/00  
CC Biosynthetic genes for spinosyn insecticide production FH  
Key Location/Qualifiers  
FT source 1. .80161  
FT /organism='Saccharopolyspora spinosa',  
Location/Qualifiers  
1. .80161  
/organism="Saccharopolyspora spinosa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:60894"

ORIGIN  
Query Match 100.0%; Score 828; DB 6; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 8.3e-117;  
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTTGCCAGGTGGCGGCAACAATCGACAGGTTGGGAGATGATGACCTGTGTCAG 60  
DB 20168 GTGTTGCCAGGTGGCGGCAACAATCGACAGGTTGGGAGATGATGACCTGTGTCAG 20227  
QY 61 CCGTTGCTGAATCGGTGCGCGGCGCGCCCTGCGGCATTCACCAAGGCTACTGGAGAAC 120  
DB 20228 CCGTTGCTGAATCGGTGCGCGGCGCGCCCTGCGGCATTCACCAAGGCTACTGGAGAAC 20287  
QY 121 GACGGGCGGGGCTTCTCTGGCAGAGGCGCGCCGACCGGCTCACCGACCTTGTGCGGAAACG 180  
DB 20288 GACGGGCGGGGCTTCTCTGGCAGAGGCGCGCCGACCGGCTCACCGACCTTGTGCGGAAACG 20347  
QY 181 ACCGTGCTGATGCGCGGCTTCTGACTGCTCGATGTGGGGTGGGTAACCGGACACACG 240  
DB 20348 ACCGTGCTGATGCGCGGCTTCTGACTGCTCGATGTGGGGTGGGTAACCGGACACACG 20407  
QY 241 CTGGCGGTCGCGCGGCAACAACCGATCAACCGATCAACCGATCAACCGATCAACCGATCAAC 300  
DB 20408 CTGGCGGTCGCGCGGCAACAACCGATCAACCGATCAACCGATCAACCGATCAACCGATCAAC 20467  
QY 301 GTGGCCATCGCGGCTGATGTCGACGCGAACCGGACTTAAGCCACCGGGTGGACTTCTCG 360  
DB 20468 GTGGCCATCGCGGCTGATGTCGACGCGAACCGGACTTAAGCCACCGGGTGGACTTCTCG 20527  
QY 361 TGGCTGATGCCATGTCCTTCCGTCGTAACCGGACCAATGCTTTGACGCGCGCTGGGCCATG 420  
DB 20528 TGGCTGATGCCATGTCCTTCCGTCGTAACCGGACCAATGCTTTGACGCGCGCTGGGCCATG 20587  
QY 421 CAGTGCATGTCGATGTCCTGCGATCCGACCGGATCCCGGGAATCTTTCGATGATCTC 480  
DB 20588 CAGTGCATGTCGATGTCCTGCGATCCGACCGGATCCCGGGAATCTTTCGATGATCTC 20647  
QY 481 AAACCCGGTGGCATCTCTGCGCTGACCGAGGTCTCAACGAGAAAGGAGGCGCGCGGATG 540  
DB 20648 AAACCCGGTGGCATCTCTGCGCTGACCGAGGTCTCAACGAGAAAGGAGGCGCGCGGATG 20707  
QY 541 CCGGTTCTCGGGGACAGGTGGCGACCGGCTTGGGATCTGCTGGCTGAGCACTTCTG 600  
DB 20708 CCGGTTCTCGGGGACAGGTGGCGACCGGCTTGGGATCTGCTGGCTGAGCACTTCTG 20767  
QY 601 GAATGCTGCGTGCAGCGGGGTTGAGATCTCTGATTTGGAGAGAGTGTGCTGAGGAGCC 660  
DB 20768 GAATGCTGCGTGCAGCGGGGTTGAGATCTCTGATTTGGAGAGAGTGTGCTGAGGAGCC 20827

QY 661 CGGTACTTCATGCCAGTTCGCGGAAGGCTCGTGCGACCAAGACGGATCGCGGAC 720  
DB 20828 CGGTACTTCATGCCAGTTCGCGGAAGGCTCGTGCGACCAAGACGGATCGCGGAC 20887  
QY 721 AGGTACGGGCGGGCTGTGCGCGGCTGGGCGCGCGGCTGTCGATTTATGAGAAATATGCC 780  
DB 20888 AGGTACGGGCGGGCTGTGCGCGGCTGGGCGCGCGGCTGTCGATTTATGAGAAATATGCC 20947  
QY 781 CACGACATGGGCTATGCGATTTCTGACGCGCGGGAAGCCGGTTCGGCTGA 828  
DB 20948 CACGACATGGGCTATGCGATTTCTGACGCGCGGGAAGCCGGTTCGGCTGA 20995

RESULT 7  
AR281866  
LOCUS AR281866 80161 bp DNA linear PAT 10-APR-2003  
DEFINITION Sequence 1 from patent US 6521406.  
ACCESSION AR281866  
VERSION AR281866.1 GI:29717767  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE Unknown.  
AUTHORS (bases 1 to 80161)  
Baltz, R.H., Broughton, M.C., Crawford, K.P., Madduri, K., Merlo, D.J.,  
Treadway, P.J., Turner, J.R., and Waldron, C.  
TITLE Sping, a gene for spinosyn insecticide biosynthesis  
JOURNAL Patent: US 6521406-A 1 18-FEB-2003;  
FEATURES  
source 1. .80161  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 100.0%; Score 828; DB 6; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 8.3e-117;  
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTTGCCAGGTGGCGGCAACAATCGACAGGTTGGGAGATGATGACCTGTGTCAG 60  
DB 20168 GTGTTGCCAGGTGGCGGCAACAATCGACAGGTTGGGAGATGATGACCTGTGTCAG 20227  
QY 61 CCGTTGCTGAATCGGTGCGCGGCGCGCCCTGCGGCATTCACCAAGGCTACTGGAGAAC 120  
DB 20228 CCGTTGCTGAATCGGTGCGCGGCGCGCCCTGCGGCATTCACCAAGGCTACTGGAGAAC 20287  
QY 121 GACGGGCGGGGCTTCTCTGGCAGAGGCGCGCCGACCGGCTCACCGACCTTGTGCGGAAACG 180  
DB 20288 GACGGGCGGGGCTTCTCTGGCAGAGGCGCGCCGACCGGCTCACCGACCTTGTGCGGAAACG 20347  
QY 181 ACCGTGCTGATGCGCGGCTTCTGACTGCTCGATGTGGGGTGGGTAACCGGACACACG 240  
DB 20348 ACCGTGCTGATGCGCGGCTTCTGACTGCTCGATGTGGGGTGGGTAACCGGACACACG 20407  
QY 241 CTGGCGGTCGCGCGGCAACAACCGATCAACCGATCAACCGATCAACCGATCAACCGATCAAC 300  
DB 20408 CTGGCGGTCGCGCGGCAACAACCGATCAACCGATCAACCGATCAACCGATCAACCGATCAAC 20467  
QY 301 GTGGCCATCGCGGCTGATGTCGACGCGAACCGGACTTAAGCCACCGGGTGGACTTCTCG 360  
DB 20468 GTGGCCATCGCGGCTGATGTCGACGCGAACCGGACTTAAGCCACCGGGTGGACTTCTCG 20527  
QY 361 TGGCTGATGCCATGTCCTTCCGTCGTAACCGGACCAATGCTTTGACGCGCGCTGGGCCATG 420  
DB 20528 TGGCTGATGCCATGTCCTTCCGTCGTAACCGGACCAATGCTTTGACGCGCGCTGGGCCATG 20587  
QY 421 CAGTGCATGTCGATGTCCTGCGATCCGACCGGATCCCGGGAATCTTTCGATGATCTC 480  
DB 20588 CAGTGCATGTCGATGTCCTGCGATCCGACCGGATCCCGGGAATCTTTCGATGATCTC 20647  
QY 481 AAACCCGGTGGCATCTCTGCGCTGACCGAGGTCTCAACGAGAAAGGAGGCGCGCGGATG 540  
DB 20648 AAACCCGGTGGCATCTCTGCGCTGACCGAGGTCTCAACGAGAAAGGAGGCGCGCGGATG 20707

```

Db      20648 AACCCTGCTGATCTCTGCGCTGCAACCGAGTCTGTAACGAGAAACGGGCGCGGGGATG 20707
Qy      541 CCGGTGTCCGGGAGACAGTGGCCGACCGGCTTTGGATTCGCTGGCTGAGCACTTCTG 600
Db      20708 CCGGTGTCCGGGAGACAGTGGCCGACCGGCTTTGGATTCGCTGGCTGAGCACTTCTG 20767
Qy      601 GAATGCTGCGGAGAGCGGGGGTTCCGATCTCGATTTGGAGAGAGTGTGCTGAGGACC 660
Db      20768 GAATGCTGCGGAGAGCGGGGGTTCCGATCTCGATTTGGAGAGAGTGTGCTGAGGACC 20827
Qy      661 CGTACTTATGCTCCGAGTTCGCGAAGAGCTGCTGCGCACCAGACCGGAGATCGCGAC 720
Db      20828 CCGTACTTATGCTCCGAGTTCGCGAAGAGCTGCTGCGCACCAGACCGGAGATCGCGAC 20887
Qy      721 AGTACGCGGCGCGCTGCTGCGCGGCTGCGCGCGCTGCTGCGATTTAGAAATATGCC 780
Db      20888 AGTACGCGGCGCGCTGCTGCGCGGCTGCGCGCGCTGCTGCGATTTAGAAATATGCC 20947
Qy      781 CACGACATGGGCTATGCGATTTGACAGCGGCGGAAACCGGTCGGCTGA 828
Db      20948 CACGACATGGGCTATGCGATTTGACAGCGGCGGAAACCGGTCGGCTGA 20995

RESULT 8
LOCUS   AX600587/c 36538 bp DNA linear PAT 14-FEB-2003
DEFINITION
Sequence 2 from Patent WO02079477.
ACCESSION
AX600587
VERSION
AX600587.1 GI:28400303
KEYWORDS
SOURCE
Saccharopolyspora sp. NRRL 30141
ORGANISM
Saccharopolyspora sp. NRRL 30141
Bacteria; Actinobacteridae; Actinomycetales;
Pseudonocardiales; Pseudonocardaceae; Saccharopolyspora.
REFERENCE
1 Hahn, D.R., Jackson, J.D., Bullard, B.S., Gustafson, G.D., Waldron, C.
and Mitchell, J.C.
Biothetic genes for butenyl-epinosyn insecticide production
Patent: WO 02079477-A 2 10-OCT-2002;
Dow Agrosciences LLC (US)
FEATURES
Location/Qualifiers
1..36538
/organism="Saccharopolyspora sp. NRRL 30141"
/mol_type="unassigned DNA"
/db_xref="taxon:221471"
/note="LW107129"
ORIGIN
Query Match 90.5%; Score 749.6; DB 6; Length 36538;
Best Local Similarity 94.1%; Pred. No. 8e-105;
Matches 779; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy      1 GGTGGCAGGTGGCGGACCAACATCGACAGGTTGGGAGATGATGACTGTGACG 60
Db      938 GTGTGGCAGGTGGCGGACCAACATCGACAGGTTGGGAGATGATGACTGTGACG 879
Qy      61 CCGTTGTGAATCTCGTTCGCGGGCGGCGCTGCGCCATCCACACCGGCTAATCGGAGAAC 120
Db      878 CCGTTGTGAATCTCGTTCGCGGGCGGCGCTGCGCCATCCACACCGGCTAATCGGAGAAC 819
Qy      121 GACGGGCGGGCTTCTTGACGACGCGCGGACCGGCTACCGAACCCTTGTCCGGAACGG 180
Db      818 GACGGGCGGGCTTCTTGACGACGCGCGGACCGGCTACCGAACCCTTGTCCGGAACGG 759
Qy      181 ACCGTGCTGATGGCGGCGTTCGACTGCTGATGTGGGTGCGGTACCGGACCAACGACG 240
Db      758 ACCGTGCTGATGGCGGCGTTCGACTGCTGATGTGGGTGCGGTACCGGACCAACGACG 699
Qy      241 CTGCGCGTTCGCGCGGACCAACGCGATCAATCAACCGGCTACCGTCAACGAGTGCAG 300
Db      698 CTGCGCGTTCGCGCGGACCAACGCGATCAATCAACCGGCTACCGTCAACGAGTGCAG 639
Qy      301 GTGGCCATGCCCGCTGATTTGCGACGCGGAACCGGACTAAGCCACCGGGTGAATCTTCG 360

```

```

Db      638 GCTGCAATGCGGTTGATTTGGGCGCGGACCGCGACTAAGCACAGAGGACTTCTCG 579
Qy      361 TCGTGTGATGCCATGTCCTTCGCTGACCGGACCAATGCTTTGACAGCCGCGGCGCAATG 420
Db      578 TCGTGTGATGCCATGTCCTTCGCTGACCGGACCAATGCTTTGACAGCCGCGGCGCAATG 519
Qy      421 CAGTGTGTTGGAGATGTCGGAACCGGACCGGTCATCCGGGAAATCTTGAATGACTC 480
Db      518 CAGTGTGTTGGAGATGTCGGAACCGGACCGGTCATCCGGGAAATCTTGAATGACTC 459
Qy      481 AACCCTGCTGATCTCGGCGCTCAACGAGTCTGTAACGAGAAACGGGCGGCGGATG 540
Db      458 AACCCTGCTGATCTCGGCGCTCAACGAGTCTGTAACGAGAAATAGCGAGCGGATG 399
Qy      541 CCGGTGTCCGGGAGACAGGTGGCGACCGGCGCTTGGATTTCTGCTGCTGAGCACTTCTG 600
Db      398 CCGGTGTCCGGGAGACAGGTGGCGACCGGCGCTTGGATTTCTGCTGCTGAGCACTTCTG 339
Qy      601 GAATGCTGCGGAGAGCGGGGTTGAGATCTCTGATGGAGAGAGAGTGTCTGAGAGACC 660
Db      338 GAATGCTGCGGAGAGCGGGGTTGAGATCTCTGATGGAGAGAGAGTGTCTGAGAGACC 279
Qy      661 CGTACTTATGCTCCGAGTTCGCGAAGAGCTGCTGCGCACACGACGCGGATCGCGAC 720
Db      278 CCGTACTTATGCTCCGAGTTCGCGAAGAGCTGCTGCGCACACGACGCGGATCGCGAC 219
Qy      721 AGTACGCGGCGCGCTGCTGCGCGGCTGCGCGCGGCTGCTGCGATTTAGAAATATGCC 780
Db      218 AGTACGCGGCGCGCTGCTGCGCGGCTGCGCGCGGCTGCTGCGATTTAGAAATATGCC 159
Qy      781 CACGACATGGGCTATGCGATTTGACAGCGGCGGAAACCGGTCGGCTGA 828
Db      158 GACGACATGGGCTATGCGATTTGACAGCGGCGGAAACCGGTCGGCTGA 111

```

```

RESULT 9
LOCUS   AX598613 846 bp DNA linear PAT 14-FEB-2003
DEFINITION
Sequence 21 from Patent WO02088176.
ACCESSION
AX598613
VERSION
AX598613.1 GI:28398761
KEYWORDS
SOURCE
Streptomyces platensis
ORGANISM
Streptomyces platensis
Bacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE
1 Farnet, C.M., Zazopoulos, E., Staffa, A. and Yang, X.
Genes and proteins for the biosynthesis of polyketides
Patent: WO 02088176-A 21 07-NOV-2002;
Ecopia Biosciences Inc. (CA)
FEATURES
Location/Qualifiers
1..846
/organism="Streptomyces platensis"
/mol_type="unassigned DNA"
/db_xref="taxon:58346"
/note="Subsp. tobacensis"
ORIGIN
Query Match 20.5%; Score 170; DB 6; Length 846;
Best Local Similarity 59.4%; Pred. No. 1.9e-16;
Matches 348; Conservative 0; Mismatches 215; Indels 9; Gaps 2;

Qy      5 TGCAGAGTGGCGGACCAACATGCGACAGGTTGGGAGATGATGACTGTGACGCGGT 64
Db      35 TCCCGCTTCGCGCGCGCGCTCCGGAAGAGTGGACACTTACGACCCCTTCACCGCAC 94
Qy      65 TGTGTAATCTCGTTCGCGGGCGGCGCTGCGCATCCACACCGGCTAATCGGAGAAACGACG 124
Db      95 TGTGTAATCTCGTTCGCGGGCGGCGCTGCGCATCCGCTGCTGATGAGAGTGTGACG 148
Qy      125 GCGGCGCTTCCTGGC---AGCAGGCGCGGACCGGCTCAACCGACTTGTGCGGAAACGGA 181

```

149 ACAACGACACCCCGCTCGTGGAGGGCCGACCGGCTCACCGACACAGATGATCCGACCGCC 208  
182 CCGTGCTCGATGGCCGCGCTTCGACTGCTCGATGATGGGGTGGCGGTACCGGACACAGAGCGC 241  
209 TGGCGATGACACGAGGACAGCGGGTCTCTGACGTGGGCTGGGAGTGGCCAGCGCGCA 268  
242 TGGCGGTGGCGCGGCACACGCGATCCAGATCACCGGCTATCACCGTCAAGCTGGTCAAG 301  
269 TGGCGATCGCGCGGCGCACCGCGCGCCATGTACAGGAGTCGCGATCAGCAAGACAGCA 328  
302 TGGCCATGGCGCGCTGATTTGGCGACCGCGAACGGGACTAAGCCACCGGGGTGACTTCTGCT 361  
329 TGGCGCGGCGCACCGCGCTCGCGAGGGGCGCGGCTGAGCCACCGCGTGGAGTTCCGGC 388  
362 GCGTGATGGCATGTCCTCCGTCGACCGGACAATGCTTTGACAGCGCGGCTGGGCGCATGC 421  
389 ACGCGGAGCCATGGAACTGCGCTTCCCGAGACTCTTTGACGCGCGCATGCGCATG 448  
422 AGTGCCTGTGGAGATGTCCGAACCGGACCGTGCATCCGGGAAATCTTTCGAGTACTCA 481  
449 AGTGATCTTCCACATGCCCCGACCGCGAGCGGGTCTCGCGCAGATCGCGCCGCTACTGC 508  
482 AACCGGTGGCATCTCGCGCTACCGAGTGTGTAACAGAGAGCGGGCGGGAGATGC 541  
509 GCCCGGGGCGCGCTGGTCTTCAACGACTTCTTGAAGCGCGGCCCGTCCCGCCGAGA 568  
542 CGGTGTCCGGGG 553  
569 AGCAACCCGCG 580

RESULT 10	AX598593				
LOCUS	AX598593	52101 bp	DNA	linear	PAT 14-FEB-2003
DEFINITION	Sequence 1 from Patent WO02088176.				
ACCESSION	AX598593				
VERSION	AX598593.1	GI:28398751			
KEYWORDS	.				
SOURCE	Streptomyces platensis				
ORGANISM	Streptomyces platensis Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
REFERENCE	1				
AUTHORS	Farnet,C.M., Zazopoulos,E., Staiffa,A. and Yang,X.				
TITLE	Genes and proteins for the biosynthesis of polyketides				
JOURNAL	Patent: WO 02088176-A 1 07-NOV-2002; Ecopia Biosciences Inc. (CA)				
FEATURES	Location/Qualifiers				
SOURCE	1..52101 /organism="Streptomyces platensis" /mol_type="unassigned DNA" /db_xref="taxon:58346" /note="subsp. rosaceus"				
ORIGIN					

	Query Match	Similarity	20.5%;	Score: 170;	DB 6;	Length 52101;
	Best Local	Similarity	59.4%;	Pred. No. 7.1e-17;		
	Matches	328;	Conservative	0;	Mismatches 215;	Indels 9;
					Gaps	2;
QY	5	TGCCAGGTGGCGCACCAACATCGCAGCGTTGGGCGAGATGTATGACCTGATCAGCCCGT	64			
DB	51268	TCCCCGTCCTCGGCCCTCCGTCGCCGAAGAAGTTCGGAACCTTACGACCGGCTCACCGCAC	51327			
QY	65	TGCTGAACCTCGGTGGCGGGCGGCCCTTGCGCCATCCACACGGGCTACTGGAGATGACG	124			
DB	51328	TGGAACACGAAGCGGCCGGCG-----GCAAGCTTCAACCTCGGCTACTGGAGATGACG	51381			
QY	125	GGCGGGCTTCTTGCG---AGACGGCGCGCGACCGGCTCAACGACCTTGTGGCGGAACGGA	181			
DB	51382	ACAACGACACCCCGCTGTGTGGAGCGGCCCAACCGGCTCAACCGACACAGATGACCGACCGCC	51441			
QY	182	CCGTGCTCGATGGCGGCGTTTCGACTGCTCGATGTGGGGTTCGGTACCGGACCAACGACGC	241			

Db 51442 TGCAGATCGACAGGAGACAGGGATCTCTGACGTCGGAGTCCGACAGCCGCCA 51501  
 QY 242 TGCAGCTGCAGCGCGACMACGCATCTCAATCATCCGACATCATCCGTACGCCAGTGCAG 301  
 Db 51502 TGCAGATCGCCCGCGCAGCAGCGGCCCATATGTCAACGGGCATCGGATCAACAAGACACA 51561  
 QY 302 TTGGCATTCGCGCGCGATGTGGCAACGGGAACGGGAGCTAAGGCCACGGGGTGAATTCTCGT 361  
 Db 51562 TCGCCCGCGCACCGCCCTTGCAGAGGCGCGGCTTGACGACGCGCTGAGATTTCGGC 51621  
 QY 362 GCGTCGATGTCATGATCTCCCTGCGGTACCCGACCAATGCTTTGACGCGCGCTGGGCCATGC 421  
 Db 51622 ACGCCAGCGCATGAACTGCTCCCTTCCCGACGACTCTTTCAGCGCGGCATTCGCATGC 51681  
 QY 422 AGTGGCTTTGGAGATATTCGGAACCGGACCGTGGCATTCGGGAAATCTTTCGAGATCTCA 481  
 Db 51682 AGTGAATCTTCCATATGCCGACCGGAGGGATCTCTGCCGAGATCTCCCGCGTACTGC 51741  
 QY 482 AACCCGCTGCGATCTTCGCGGTCAACCGAGGTCTGTCAACGAGAACGGGCGCGCGGATATGC 541  
 Db 51742 GCCCGGCGGCGCGCTGATCTCAACGACTTTTTCGAGGCGGCGCCGATCTCCGCGCAGGA 51801  
 QY 542 CGGTGTCGGGG 553  
 Db 51802 AGCAGCCGCGG 51813

RESUT 11					
AX089434					
LOCUS	AX089434	900 bp	DNA	linear	PAT 21-MAR-2001
DEFINITION	Sequence 19 from Patent WO0116303.				
ACCESSION	AX089434				
VERSION	AX089434.1	GI:13443695			
KEYWORDS	.				
SOURCE					
ORGANISM	<i>Saccharopolyspora spinosa</i> <i>Saccharopolyspora spinosa</i> Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardiales; Saccharopolyspora.				
REFERENCE	1 Ebertz, G., Moehrlie, V., Froede, R., Velten, R. and Salas, J. A. Nucleic acids which code for the enzyme activities of the spinosyn biosynthesis Patent: WO 0116303-A 19 08-MAR-2001;				
AUTHORS	BAYER AG (DE)				
TITLE	JOURNAL				

FEATURES	SOURCE	Location/Qualifiers
		1..900
		/organism="Saccharopolyspora spinosa"
		/mol_type="unassigned DNA"
		/db_xref="taxon:60894"
	CDS	1..900
		/note="unnamed protein product; ORF7; O-Methyltransferase"
		/codon_start=1
		/transl_table=1
		/protein_id="CAC34880.1"
		/db_xref="GI:13443696"
		/translation="MNSRGANTQAFPTADYVESIFDALAGRPLHNGYMAQGRPDAG ATPESDADQDTLDFIDKALRRGAHLFDJGCGNGQPVRAACASGVRTGITYNAOH LAAATRLANSTGSLSEFPLVDGAQLPTGDFEQAMAMQSVQIYDQAAALREVRH ILIEGRFVLEGLDITTRVRLPEEYAAVWTGPAHLNLFALVBSAGEIILEVDTLQ TRCVSNVYVDELLRKDELGAQVEPAAVGTQGRYLDIAAKHGGPRLLAALVAEYRK HPDYARBSMSGFWLLQARKKQS"
ORIGIN		
Query Match	20.0%; Score 165.4; DB 6; Length 900;	
Best Local Similarity	52.8%; Pred. No. 9,2e-16;	
Matches 386; Conservative	0; Mismatches 336; Indels	9; Gaps 1;
QY	15	CGACCAACATCGACGAGGTTGGGACAGATGTATGACTTGTCACGCCGTTGTGAATTC 74
DB	21	CACACAGGCAATTTCCGACCGCGGATCAGGTGAGATCCATCTTCATGATGCGTGGGCAACGG 80
QY	75	GATGCGGGGCGCCCTGCGCCATCCACCAACGCGTACTGGAGAACGACGGCGGGCTTC 134

[illegible][illegible]



Db 1893 CCAACGGCTCACCGCATCTGTGCGCCGAGCGCTCGCGCTCCGCCCGACCGGCATCTGCT 1952  
QY 210 CGATGCGGGGCTGGATACCGGACCAACCAAGCGTGGCGCGCGCGGACCAAGCCATCCA 269  
Db 1953 GGAAGTGGGCTGGGATGGCGTGGCGCTCTCAAGATGGCGGAGCCGACGACGTCCG 2012  
QY 270 GATCAACCGGCATCACCGTCAAGCCAGGTGAGTGGCCATCGCGCTGATTGGCAGCGA 329  
Db 2013 CCGTCAACCGGGATCACCGTCAAGCCAGGATGACCGAGGGCGCGAGCGGGCGGTGGA 2072  
QY 330 ACCCGGACTTAACCAACCGGGTGAATCTTCGTGCGCTGATGCCATGTCCTCCGTAACC 389  
Db 2073 GTCCATGCGGGGCGGGGCTCTCTTCGCGCTGGGAGCGCCATGAACTCTCCCTTCCA 2132  
QY 390 GGACATGCTTTCAGACGCGCGCTGGGGCAATGAGTGCCTGTTGAGATGTCCGAACCGGA 449  
Db 2133 GAACTCTCTTTCAGACGCGCGCTTGGCCATGAGTGGCTGTGATCTGCCGACGAC 2192  
QY 450 CCGTGCATCCGGAATCTTTCAGTACTCAACCCCGGTGCGATCTTCGCGCTCAGCA 509  
Db 2193 ACCCGGCTCAAGAGATCACCGGGTGTCCGCCCGCGCGCGCGCTGTGATGCCGA 2252  
QY 510 GGTCTCAACGAGAACCGCGCGCGG 536  
Db 2253 CCGTGTGACGACGCGCTTCAACCGG 2279

RESULT 14  
AX089419  
LOCUS AX089419 45624 bp DNA linear PAT 21-MAR-2001  
DEFINITION Sequence 4 from Patent WO0116303.  
ACCESSION AX089419  
VERSION AX089419.1 GI:13443680  
KEYWORDS  
SOURCE Saccharopolyspora spinosa  
ORGANISM Saccharopolyspora spinosa  
Bacteria; Actinobacteria; Actinomycetales;  
Pseudonocardiales; Pseudonocardiales; Saccharopolyspora.

REFERENCE  
AUTHORS Eberz, G., Moehrl, V., Froede, R., Velten, R. and Salas, J.A.  
TITLE Nucleic acids which code for the enzyme activities of the spinosyn  
biosynthesis  
PATENT: WO 0116303-A 4 08-MAR-2001;  
JOURNAL BAYER AG (DE)  
FEATURES  
source 1..45624  
/organism="Saccharopolyspora spinosa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:60894"

ORIGIN  
Query Match 20.0%; Score 165.4; DB 6; Length 45624;  
Best Local Similarity 52.8%; Pred. No. 3.7e-16;  
Matches 386; Conservative 0; Mismatches 336; Indels 9; Gaps 1;  
Db 15 CGCAACAATCGACGAGTTGGCAGATGATGACCTGTGACGCGGTGCTGAATC 74  
QY 7421 CACACAGGATTTCCGACCGGAGATCAGTGAATCTTTCATGCTGGTGGCCAGG 7480  
Db 75 GGTGCGGGGCGGCTTGGCCATCACACGCGCTACTGGAGAAAGACGCGGGGCTTC 134  
QY 7481 GCGTCCCTGACACGCTTACTGGGCGGCGGTATCGGAGATGCGCGGTGACAC 7540  
QY 135 CTGGCAGAGCGCGCGACCGGCTGACCGACCTTGTCCCGCAACGAGCGGTGATGG 194  
Db 7541 GTGTGCGATGTGCGCGACCACTGATCGACCTGTTCAACGAAAGCGCGCTCGTC 7600  
QY 195 CGGCGTTGACTGCTGATGTGGGTGGGTGAGTACCGGACCAACGAGCGCTGCGCGG 254  
Db 7601 CGGAGCGGACCTGTTGACCTGGGCTGGGAGATGGGAGCGCCGTAATCGGTGGGATG 7660  
QY 255 CGACAGCGGATCCAGATCACCGGATCACCGTCAAGCGGTGCAATGGTCCATCGCGC 314

Db 7661 CGCCAGCGCGCTTGAGTACACCGGAATACCGTGAAGCCAGCATCTCGCGCGGCAC 7720  
QY 315 TGATTGCGGACCGCAACCGGACCTTAAGCACCGGGTGAATCTTCTGCTGATGCAAT 374  
Db 7721 CAGGCTTCACCAACGAGACCGGACCTGGCGGAGTCTTGAATGATTAATGACGGGCG 7780  
QY 375 GTCCCTGCGGATACCGGACCAATGCTTTCAGACGCGCGCTGGGCAATGACGCTGTTGA 434  
Db 7781 CCAAGTGGCTTAACCGGACGCTTCTTCAAGCGCGCATGGCGATGCAATCGCTGTCGA 7840  
QY 435 GATGTCGAACCGGACCGTGCATCCGGGAAATCTTTCAGTACTCAAAACCGGTGAT 494  
Db 7841 GATGTCGAACCGGACCGGCTGATCCGGATCCGGAGTCCAGCAATCTGGAAACCGGCGCG 7900  
QY 495 CCGTGGCGCTCAACGAGTCTCAAAACGAGACCGGCGCGGATGCCGTGTCGGGGA 554  
Db 7901 GTTGTCTTGGAGACATCATCTCGGTTTCGACTCCCGGAAAGTACGGCGGTTTG 7960  
QY 555 CAGTGGCGGACCGGCGCTTGGATCTGCGCTGAGCACTTTCGAAATCGCTGCTGC 614  
Db 7961 GACGGGACGACCGCCCATACCTT-----GAAAGCTTCAAGCGCGCTGTGACGGA 8011  
QY 615 AGCGGAGTTGAGATCTCTGATGGAGAGACGTGCTGAGAGACCGGTAATTATGCC 674  
Db 8012 AGCGGAGTTGAGATCTCTGAAATCAACGACCTCAAGGACAGACAGATGATGCTTC 8071  
QY 675 GCAATGCGCGGAAGCTCGCTGCGGACACGACAGGATGCGGACAGTACGGGCGGCG 734  
Db 8072 CTGTGATCTGACGAGTGTCTCCGAAATCTGATGAGCTGCGCGCTGAGCTTGCGCG 8131  
QY 735 TGTGCGCGGCT 745  
Db 8132 TGTGCGGACCT 8142

RESULT 15  
AX089416/C  
LOCUS AX089416 50000 bp DNA linear PAT 21-MAR-2001  
DEFINITION Sequence 1 from Patent WO0116303.  
ACCESSION AX089416  
VERSION AX089416.1 GI:13443677  
KEYWORDS  
SOURCE Saccharopolyspora spinosa  
ORGANISM Saccharopolyspora spinosa  
Bacteria; Actinobacteria; Actinomycetales;  
Pseudonocardiales; Pseudonocardiales; Saccharopolyspora.

REFERENCE  
AUTHORS Eberz, G., Moehrl, V., Froede, R., Velten, R. and Salas, J.A.  
TITLE Nucleic acids which code for the enzyme activities of the spinosyn  
biosynthesis  
PATENT: WO 0116303-A 1 08-MAR-2001;  
JOURNAL BAYER AG (DE)  
FEATURES  
source 1..50000  
/organism="Saccharopolyspora spinosa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:60894"

ORIGIN  
Query Match 20.0%; Score 165.4; DB 6; Length 50000;  
Best Local Similarity 52.8%; Pred. No. 3.6e-16;  
Matches 386; Conservative 0; Mismatches 336; Indels 9; Gaps 1;  
QY 15 CGCAACAATCGACGAGTTGGGAGATGATGACCTGTGACGCGGTGCTGAATC 74  
Db 38204 CACACAGGATTTCCGACCGGATCAGTGAATCTTTCATGCTGGTGGCCAGG 38145  
QY 75 GGTGCGGGGCGGCTTGGCCATCACACGCGCTACTGGAGAAAGACGCGCGGCTTC 134  
Db 38144 GCGTCCCTGACACGCTTACTGGGCGGCGGTATCGGAGATGCCGCTGACAC 38085  
QY 135 CTGGCAGAGCGCGCGACCGGCTGACCGACTTGTGCGGAAAGACGCTGCTGATGG 194



```
Db      38084 GTGTCGAGTGTGCTGCCGACCAACTGACCGCACTCTGTTTCATCGACAAAGCCGCGCTCCGTCC 38025
QY      195 CCGCGTTTCGACTGCTCGATGTGGGGTGCGGTACCGGACAACCAAGCGCTGCGCGTCCGCGC 254
Db      38024 CGGAGGGACCTGTTTGAACCTGGGGCTGCCGCAATGGGCAAGCCGGTAGTCCGTGCGGCATG 37965
QY      255 CGACAAACCGGATCCAGATACCGGGCATCACCGTCAAGCCAGTGCMAAGTGGCCATGCCCGC 314
Db      37964 CCGCAGCGCGGCTTCGAGTCAACGGGAATCACCGTGAACGCCCAAGCATCTGCGCGCCGAC 37905
QY      315 TGAATTGGCGACGCGGAACCGGACCTAAGCAACCGGCGTGAATTCTCGTGCATGCCAT 374
Db      37904 CAGGCTTCGCAACGACACCGAGCTGGCCGGCACTTGAAGTTCGATCTAGTCGACGCGC 37845
QY      375 GTCCCTGCGGTACCCCGACAATGCTTTCGACGCGCGCTGGCCCATGCAAGTCGCTGTGGA 434
Db      37844 CCAGCTGGCTTACCGGACGGTTCTTTTCAAGGCGCGCATGGGCGATCGAGTCCGTGCA 37785
QY      435 GATGTCGGAACCGGACCGTCCCATCCGGGAAATCCTTCGAGTACTCAAAACCGGTGGCAT 494
Db      37784 GATCGTGAACCAAGCCGCGCGCATTCGCGAGGTCCACCGAATCTTGAAACCGGCGGCGC 37725
QY      495 CCTCGGCGTCAACGAGTGTCAACGAGAAGCGGCGGCGGATGCCGATGCCGAGGA 554
Db      37724 GTTCGTCTCGGACATCATCTCGGTTTCGATCTCCGAAAGTACCGCGCGTTTG 37665
QY      555 CAGGTGGCCGACCGGCTTCGATCTGCTGCTGCTGACCACTTCTGAAATCGCTGCGTGC 614
Db      37664 GACGGGCAAGACCGCCCATACCTT-----GAACAGTTTCAAGCGCGTGGTCAAGCA 37614
QY      615 AGCGGGGTTTCGAGATCTTCGATTGGAGAGACGTGTGTCGAGGACCGCGTACTTCATGCC 674
Db      37613 AGCGGGGTTTCGAGATCTTCGAGATTCACCGACTCACGSCACAGACCAGGTGCAATGCTTC 37554
QY      675 GCAGTTCGCGCAAGAGCTCGCTGCGGACCAAGCAAGGATCGCGGACAGGTACGCGCGCGC 734
Db      37553 CTGGTACGTGACGAGTGTCTCCGAAACTCGATGAGCTGCGCGCTCGAGCTCGGCGC 37494
QY      735 TGTCCCGGCT 745
Db      37493 TGTCCGCACCT 37483
```

Search completed: August 29, 2005, 16:10:54  
Job time : 3894 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 29, 2005, 12:38:52 ; Search time 585 Seconds

(without alignments)  
8378.711 Million cell updates/sec

Title: US-10-069-353A-7

Perfect score: 828  
Sequence: 1 gctgtccagcagtggtgcgcacc.....cgcggaagccggtcgcgcga 828

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	828	100.0	828	4	AAf88318	AAf88318 S. spinos
2	828	100.0	45624	4	AAf88315	AAf88315 S. spinos
3	828	100.0	50000	4	AAf88312	AAf88312 S. spinos
4	828	100.0	80161	2	AAz21501	AAz21501 DNA fragm
5	749.6	90.5	36538	10	ABv75558	ABv75558 Saccharop
6	170	20.5	846	10	AAz54227	AAz54227 Streptomy
7	170	20.5	52101	2	AAz54217	AAz54217 Streptomy
8	165.4	20.0	900	4	AAf88324	AAf88324 S. spinos
9	165.4	20.0	6085	2	AAf70153	AAf70153 S. longisp
10	165.4	20.0	45624	4	AAf88315	AAf88315 S. spinos
11	165.4	20.0	50000	4	AAf88312	AAf88312 S. spinos
12	165.4	20.0	80161	2	AAz21501	AAz21501 DNA fragm
13	159	19.2	36538	10	ABv75558	ABv75558 Saccharop
14	141.8	17.1	25681	10	ADf53165	ADf53165 Saccharop
15	133.2	16.1	852	3	AAf58802	AAf58802 S. lavend
16	133.2	16.1	852	10	ADf510221	ADf510221 S. lavend
17	133.2	16.1	53500	3	AAf58802	AAf58802 S. lavend
18	133.2	16.1	53500	10	ADf510221	ADf510221 S. lavend
19	130	15.7	1410	2	AAQ80524	AAQ80524 Oxidoredu
20	122.8	14.8	2122	2	AAf70152	AAf70152 S. longisp

C	21	120.2	14.5	103599	4	ABX04971
C	22	112.2	13.6	12381	3	AAZ58381
C	23	93.2	11.3	1092	12	ADf45320
C	24	89.4	10.8	1418	12	ADf45598
C	25	84	10.1	1217	13	ADf459291
C	26	83.2	10.0	828	3	AAf58801
C	27	83.2	10.0	828	10	ADf510220
C	28	74	8.9	942	10	ADf526975
C	29	71.2	8.6	795	13	ADf560017
C	30	71.2	8.6	110000	4	AAf96882_41
C	31	71.2	8.6	110000	4	AAf96883_41
C	32	67.6	8.2	67251	10	ADf26995
C	33	67.2	8.1	2000	8	ADf71938
C	34	64	7.7	2814	3	AAZ50960
C	35	63.2	7.6	1369	10	ABX12505
C	36	61	7.4	1131	11	ABD03718
C	37	61	7.4	1215	11	ABD03991
C	38	61	7.4	1215	11	ABD03909
C	39	60.6	7.3	7205	13	ADf84263
C	40	59	7.1	1862	4	AAH44261
C	41	59	7.1	2742	8	AAf61191
C	42	59	7.1	82746	8	AAf61324
C	43	58.4	7.1	2757	8	AAf38157
C	44	57.2	6.9	991	10	ABX12504
C	45	57.2	6.9	1152	10	ADf65628

## ALIGNMENTS

RESULT 1	AAf88318	standard; DNA; 828 BP.
ID	AAf88318	
XX	AAf88318	
XX	28-AUG-2001 (first entry)	
XX	S. spinosa DNA fragment encoding ORF1, SEQ ID 7.	
XX	Forosamine; trimethylrhannose; polyketide synthase; biosynthesis;	
XX	KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;	
XX	KM macrolide; insecticidal; O-methyltransferase; de.	
XX	Saccharopolyspora spinosa.	
XX	DE19957268-Al.	
XX	08-MAR-2001.	
XX	29-NOV-1999; 99DE-01057268.	
XX	27-AUG-1999; 99DE-01040596.	
XX	(FARB) BAYER AG.	
XX	Eberz G, Moehle V, Froede R, Velten R, Salas JA;	
XX	WPI; 2001-267102/28.	
XX	P-PSDB; AAB70948.	
XX	New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for	
XX	recombinant production of insecticidal spinosyns and their derivatives.	
XX	Claim 7a; Page 102-104; 354pp; German.	
XX	This invention describes a novel method nucleic acid (i) and its encoded	
XX	polypeptide (ii) containing at least one region that encodes an enzymatic	
XX	activity involved in biosynthesis of spinosyns. (i) are used (i) to	
XX	identify, inactivate or modulate genes involved in the biosynthesis of	
XX	(ii); (ii) to generate a library of polyketide synthases; (iii) for	
XX	adding forosamine or trimethylrhannose to a spinosyn or polyketide	
XX	aglycone; and (iv) for recombinant production of the corresponding	

ABX04971 S. cinam  
AAZ58381 Streptomy  
ADf45320 Rice isop  
ADf45598 Wheat iso  
ADf459291 Bacterial  
AAf58801 S. lavend  
ADf510220 S. lavend  
ADf526975 Sorangium  
ADf60017 Bacterial  
Continuation (42 o  
Continuation (42 o  
ADf26995 Sorangium  
ADf71938 Rice gene  
AAZ50960 A. haloph  
ABX12505 CDNA enco  
ABD03718 Pseudomon  
ABD03991 Pseudomon  
ABD03909 Pseudomon  
ADf84263 Aspergill  
AAH44261 Physcomit  
AAf61191 Actinosyn  
AAf61324 Actinosyn  
AAf38157 Prokaryot  
ABX12504 CDNA enco  
ADf65628 S. cerevis

enzymes, which are used for production of (II), their precursors or derivatives, including production of transgenic plants that express (II) and thus have increased resistance to insects. (I) are also useful as markers for sequencing of the *Saccharopolyspora spinosa* genome. (II) are macrolides with insecticidal, but not antibacterial, activity, and can also be used to raise specific antibodies, useful for identifying expression clones in a gene bank. Cells transformed with (I) may produce CC (II) at significantly increased levels or produce new derivatives of CC (II). This sequence, ORF 1, encodes an *S. spinosa* O-methyltransferase protein

XX Sequence 828 BP; 140 A; 262 C; 285 G; 141 T; 0 U; 0 Other;

Query Match 100.0%; Score 828; DB 4; Length 828;

Best Local Similarity 100.0%; Pred. No. 6.2e-168;

Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GTGTTGCCAGGTGGCGCACCAATCGACAGGTTGGGCAATGTAATCTGTCACG 60
DB 1 GTGTTGCCAGGTGGCGCACCAATCGACAGGTTGGGCAATGTAATCTGTCACG 60
QY 61 CGTTGCTGAACCTGGTGGCGGCGGCGCTGCGCATCACACGGGCTACCTGGGAGAC 120
DB 61 CGTTGCTGAACCTGGTGGCGGCGGCGCTGCGCATCACACGGGCTACCTGGGAGAC 120
QY 121 GACGGGCGGGCTTCTTGCGACGAGCGCGGCGGCTGCGCATCACACGGGCTACCTGGGAGAC 180
DB 121 GACGGGCGGGCTTCTTGCGACGAGCGCGGCGGCTGCGCATCACACGGGCTACCTGGGAGAC 180
QY 181 ACCGTCCTGATGCGCGGCGGCTGCTGATGTTGGGGTGGGTTACCGGACCAACGAGG 240
DB 181 ACCGTCCTGATGCGCGGCGGCTGCTGATGTTGGGGTGGGTTACCGGACCAACGAGG 240
QY 241 CTGGCGGTGCGCGGCGGACCAACCGGATCCAGATCACCGGCGGATCACCGTCCAGGAGCA 300
DB 241 CTGGCGGTGCGCGGCGGACCAACCGGATCCAGATCACCGGCGGATCACCGTCCAGGAGCA 300
QY 301 GTGGCCATGCGCGCTGATTTGCGACGCGGACCGGAGCTAAGCCACCGGGTGGACTTCTCG 360
DB 301 GTGGCCATGCGCGCTGATTTGCGACGCGGACCGGAGCTAAGCCACCGGGTGGACTTCTCG 360
QY 361 TGGCGTCGATGCTATCTCTGCGGTTACCGGACCAATGCTTTGAGAGCGCGCTGGGCGATG 420
DB 361 TGGCGTCGATGCTATCTCTGCGGTTACCGGACCAATGCTTTGAGAGCGCGCTGGGCGATG 420
QY 421 CAGTCGCTGTTGGAAGTGTCCGAACCGGACCGGTCATCCGGGAAATCTTTCGAGTACTC 480
DB 421 CAGTCGCTGTTGGAAGTGTCCGAACCGGACCGGTCATCCGGGAAATCTTTCGAGTACTC 480
QY 481 AAACCCGGTGCGATCTCGGCGTCAACCGAGTCTCAAAAGAAAGCGGCGCGGAGTGC 540
DB 481 AAACCCGGTGCGATCTCGGCGTCAACCGAGTCTCAAAAGAAAGCGGCGCGGAGTGC 540
QY 541 CCGGTGTCTGGGGAAGAGTGGCGGACCGGCTTCCGATTTGCTGCGTGAAGCACTTCTG 600
DB 541 CCGGTGTCTGGGGAAGAGTGGCGGACCGGCTTCCGATTTGCTGCGTGAAGCACTTCTG 600
QY 601 GAATGGCTGCGTGAAGCGGCGGTTTCAGATCTTCGATTTGGGAGAGAGTGTGCGAGGAC 660
DB 601 GAATGGCTGCGTGAAGCGGCGGTTTCAGATCTTCGATTTGGGAGAGAGTGTGCGAGGAC 660
QY 661 CGGTACTTCAATCCGAGTTCGCGAAGAGTGTGCGGACCAACGAGGATTCGCGAGC 720
DB 661 CGGTACTTCAATCCGAGTTCGCGAAGAGTGTGCGGACCAACGAGGATTCGCGAGC 720
QY 721 AGGTACGGGCGCGGCTGTGCGCGGCGGCGCGCGGCTGCGATTTAGAAATATATGCC 780
DB 721 AGGTACGGGCGCGGCTGTGCGCGGCGGCGCGCGGCTGCGATTTAGAAATATATGCC 780
QY 781 CACGACATGAGGCTATGCGATTTGACAGGCGCGGAAAGCGGCTGCA 828
DB 781 CACGACATGAGGCTATGCGATTTGACAGGCGCGGAAAGCGGCTGCA 828

```

# RESULT 2

AAAF8315/C

ID AAAF8315 standard; DNA; 45624 BP.

XX AAAF8315;

XX 28-AUG-2001 (first entry)

XX S. spinosa DNA fragment SEQ ID 4.

XX Forosamine; trimethylrharnnose; polyketide synthase; biosynthesis;

XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;

XX macrolide; insecticidal; ds.

OS Saccharopolyspora spinosa.

XX DEL19957268-A1.

XX 08-MAR-2001.

XX 29-NOV-1999; 99DE-01057268.

XX 27-AUG-1999; 99DE-01040596.

XX (FARB ) BAYER AG.

XX Ebertz G, Moehrl V, Froede R, Velten R, Salas JA;

XX WPI; 2001-267102/28.

XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for recombinant production of insecticidal spinosyns and their derivatives.

XX Claim 7; Page 58-74; 354pp; German.

CC This invention describes a novel method nucleic acid (I) and its encoded polypeptide (II) containing at least one region that encodes an enzymatic activity involved in biosynthesis of spinosyns. (I) are used (i) to identify, inactivate or modulate genes involved in the biosynthesis of (II); (ii) to generate a library of polyketide synthases; (iii) for adding forosamine or trimethylrharnnose to a spinosyn or polyketide aglycone; and (iv) for recombinant production of the corresponding enzymes, which are used for production of (II), their precursors or derivatives, including production of transgenic plants that express (II) and thus have increased resistance to insects. (I) are also useful as markers for sequencing of the *Saccharopolyspora spinosa* genome. (II) are macrolides with insecticidal, but not antibacterial, activity, and can also be used to raise specific antibodies, useful for identifying expression clones in a gene bank. Cells transformed with (I) may produce CC (II) at significantly increased levels or produce new derivatives of CC (II). This sequence represents a genomic DNA fragment of the *S. spinosa* genome which contains the coding regions for proteins involved in forosamine and trimethylrharnnose biosynthesis

XX Sequence 45624 BP; 7933 A; 14913 C; 14940 G; 7838 T; 0 U; 0 Other;

Query Match 100.0%; Score 828; DB 4; Length 45624;

Best Local Similarity 100.0%; Pred. No. 7.8e-168;

Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GTGTTGCCAGGTGGCGCACCAATCGACAGGTTGGGCAATGTAATCTGTCACG 60
DB 1 GTGTTGCCAGGTGGCGCACCAATCGACAGGTTGGGCAATGTAATCTGTCACG 60
QY 61 CGTTGCTGAACCTGGTGGCGGCGGCGCTGCGCATCACACGGGCTACCTGGGAGAC 120
DB 61 CGTTGCTGAACCTGGTGGCGGCGGCGCTGCGCATCACACGGGCTACCTGGGAGAC 120
QY 121 GACGGGCGGGCTTCTTGCGACGAGCGCGGCGGCTGCGCATCACACGGGCTACCTGGGAGAC 180
DB 121 GACGGGCGGGCTTCTTGCGACGAGCGCGGCGGCTGCGCATCACACGGGCTACCTGGGAGAC 180
QY 181 ACCGTCCTGATGCGCGGCGGCTGCTGATGTTGGGGTGGGTTACCGGACCAACGAGG 240
DB 181 ACCGTCCTGATGCGCGGCGGCTGCTGATGTTGGGGTGGGTTACCGGACCAACGAGG 240
QY 241 CTGGCGGTGCGCGGCGGACCAACCGGATCCAGATCACCGGCGGATCACCGTCCAGGAGCA 300
DB 241 CTGGCGGTGCGCGGCGGACCAACCGGATCCAGATCACCGGCGGATCACCGTCCAGGAGCA 300
QY 301 GTGGCCATGCGCGCTGATTTGCGACGCGGACCGGAGCTAAGCCACCGGGTGGACTTCTCG 360
DB 301 GTGGCCATGCGCGCTGATTTGCGACGCGGACCGGAGCTAAGCCACCGGGTGGACTTCTCG 360
QY 361 TGGCGTCGATGCTATCTCTGCGGTTACCGGACCAATGCTTTGAGAGCGCGCTGGGCGATG 420
DB 361 TGGCGTCGATGCTATCTCTGCGGTTACCGGACCAATGCTTTGAGAGCGCGCTGGGCGATG 420
QY 421 CAGTCGCTGTTGGAAGTGTCCGAACCGGACCGGTCATCCGGGAAATCTTTCGAGTACTC 480
DB 421 CAGTCGCTGTTGGAAGTGTCCGAACCGGACCGGTCATCCGGGAAATCTTTCGAGTACTC 480
QY 481 AAACCCGGTGCGATCTCGGCGTCAACCGAGTCTCAAAAGAAAGCGGCGCGGAGTGC 540
DB 481 AAACCCGGTGCGATCTCGGCGTCAACCGAGTCTCAAAAGAAAGCGGCGCGGAGTGC 540
QY 541 CCGGTGTCTGGGGAAGAGTGGCGGACCGGCTTCCGATTTGCTGCGTGAAGCACTTCTG 600
DB 541 CCGGTGTCTGGGGAAGAGTGGCGGACCGGCTTCCGATTTGCTGCGTGAAGCACTTCTG 600
QY 601 GAATGGCTGCGTGAAGCGGCGGTTTCAGATCTTCGATTTGGGAGAGAGTGTGCGAGGAC 660
DB 601 GAATGGCTGCGTGAAGCGGCGGTTTCAGATCTTCGATTTGGGAGAGAGTGTGCGAGGAC 660
QY 661 CGGTACTTCAATCCGAGTTCGCGAAGAGTGTGCGGACCAACGAGGATTCGCGAGC 720
DB 661 CGGTACTTCAATCCGAGTTCGCGAAGAGTGTGCGGACCAACGAGGATTCGCGAGC 720
QY 721 AGGTACGGGCGCGGCTGTGCGCGGCGGCGCGCGGCTGCGATTTAGAAATATATGCC 780
DB 721 AGGTACGGGCGCGGCTGTGCGCGGCGGCGCGCGGCTGCGATTTAGAAATATATGCC 780
QY 781 CACGACATGAGGCTATGCGATTTGACAGGCGCGGAAAGCGGCTGCA 828
DB 781 CACGACATGAGGCTATGCGATTTGACAGGCGCGGAAAGCGGCTGCA 828

```



[illegible]

```

FT      /tag= ac
FT      /note= "hnoyl reductase domain (BR2) : part of extender
FT      module 2"
FT      34082..34621
FT      /tag= ad
FT      /note= "Beta-ketoreductase domain (KR2) : part of extender
FT      module 2"
FT      34886..30295
FT      /tag= ae
FT      /note= "Acyl carrier protein domain (ACP2) : part of
FT      extender module 2"
FT      35419..44931
FT      /tag= af
FT      /product= "spnc"
FT      /note= "Spinosyn biosynthesis protein C"
FT      35518..36786
FT      /tag= ag
FT      /note= "Beta-ketosynthase domain (KS3) : part of extender
FT      module 3"
FT      37108..38097
FT      /tag= ah
FT      /note= "Acyl transferase domain (AT3) : part of extender
FT      module 3"
FT      38916..35374
FT      /tag= y
FT      /product= "spnc"
FT      /note= "Spinosyn biosynthesis protein C"
FT      38992..39528
FT      /tag= ai
FT      /note= "Beta-ketoreductase domain (KR3) : part of extender
FT      module 3"
FT      39790..40035
FT      /tag= aj
FT      /note= "Acyl carrier protein domain (ACP3) : part of
FT      extender module 3"
FT      40102..41373
FT      /tag= ak
FT      /note= "Beta-ketosynthase domain (KS4) : part of extender
FT      module 4"
FT      41713..42705
FT      /tag= al
FT      /note= "Acyl transferase domain (AT4) : part of extender
FT      module 4"
FT      43615..44157
FT      /tag= am
FT      /note= "Beta-ketoreductase domain (KR4) : part of extender
FT      module 4"
FT      44431..44676
FT      /tag= an
FT      /note= "Acyl carrier protein domain (ACP4) : part of
FT      extender module 4"
FT      44966..59752
FT      /tag= ao
FT      /product= "spnd"
FT      /note= "Spinosyn biosynthesis protein D"
FT      45077..46348
FT      /tag= ap
FT      /note= "Beta-ketosynthase domain (KS5) : part of extender
FT      module 5"
FT      46691..47674
FT      /tag= aq
FT      /note= "Acyl transferase domain (AT5) : part of extender
FT      module 5"
FT      47753..48310
FT      /tag= ar
FT      /note= "Dehydratase domain (DH5) : part of extender module
FT      5"
FT      49226..49771
FT      /tag= as
FT      /note= "Beta-ketoreductase domain (KR5) : part of extender
FT      module 5"
FT      50009..50254
FT      misc_feature

```

```

Query Match      100.0%; Score 828; DB 2; Length 80161;
Best Local Similarity 100.0%; Pred. No. 8,1e-168;
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GTGTTGCCAGTGTGGCGCACCAATCGACGAGTTGGGAGATGTATGACTGTGTACG 60
DB      20168 GTGTTGCCAGTGTGGCGCACCAATCGACGAGTTGGGAGATGTATGACTGTGTACG 20227

QY      61 CCGTTGCTGAACCTCGGTGCGGGCGGGCCCTTCGCCCATTCACACAGCGCTATCGGAGAAC 120
DB      20228 CCGTTGCTGAACCTCGGTGCGGGCGGGCCCTTCGCCCATTCACACAGCGCTATCGGAGAAC 20287

QY      121 GACGGGCGGGCTTCCCTGGCAGCAGGCGGCCGACCGGCTCAACGACCTTGTCCGGAACG 180
DB      20288 GACGGGCGGGCTTCCCTGGCAGCAGGCGGCCGACCGGCTCAACGACCTTGTCCGGAACG 20347

QY      181 ACCGTGCTGATGATGCGGCGCTTGACTGTGATGTGTGGGGTGTGGTACCGGACACGACG 240
DB      20348 ACCGTGCTGATGATGCGGCGCTTGACTGTGATGTGTGGGGTGTGGTACCGGACACGACG 20407

QY      241 CTGCGCGTGGCGCGCGACCAACCGCATCCAGATCAACCGCATCAACCGCATCAACCGCAT 300
DB      20408 CTGCGCGTGGCGCGCGACCAACCGCATCCAGATCAACCGCATCAACCGCATCAACCGCAT 20467

QY      301 GTGGCCATGCGCGCTGATTTGGCGGACGGAACGCGGACTAAGCCACCGGGTGAATTCTCG 360
DB      20468 GTGGCCATGCGCGCTGATTTGGCGGACGGAACGCGGACTAAGCCACCGGGTGAATTCTCG 20527

QY      361 TGGGTGATGCCATGTCTCTGCGGTACCCGGAACAATCTTTGACGCGCGCTGGGCGCATG 420
DB      20528 TGGGTGATGCCATGTCTCTGCGGTACCCGGAACAATCTTTGACGCGCGCTGGGCGCATG 20587

QY      421 CAGTCCCTGTTGAGATGTTCGCCAACCGGACCGTGCATCCGGGAAATCTTCGAGTAATC 480
DB      20588 CAGTCCCTGTTGAGATGTTCGCCAACCGGACCGTGCATCCGGGAAATCTTCGAGTAATC 20647

QY      481 AAACCCGGTGGACATCTCTGCGCGGTACCGAGGTGTGTCAAAAGAAACGGGCGCGGGATG 540
DB      20648 AAACCCGGTGGACATCTCTGCGCGGTACCGAGGTGTGTCAAAAGAAACGGGCGCGGGATG 20707

QY      541 CCGGTGTCCGGGAGCAGTGTGCGGACCGGCGCTTTGGATCTGCTGCTGAGCAATTTCTG 600
DB      20708 CCGGTGTCCGGGAGCAGTGTGCGGACCGGCGCTTTGGATCTGCTGCTGAGCAATTTCTG 20767

QY      601 GAATCCGTGCTGACGCGGGGTTTCGAGATCTTGATTTGGGAGGACCTGTCTGAGAGACC 660
DB      20768 GAATCCGTGCTGACGCGGGGTTTCGAGATCTTGATTTGGGAGGACCTGTCTGAGAGACC 20827

QY      661 CGGTACTTCAATGCGGAGTTGCGGAAGAGCTGCTGCGCACGACGACGGGATCGCGGAC 720
DB      20828 CGGTACTTCAATGCGGAGTTGCGGAAGAGCTGCTGCGCACGACGACGGGATCGCGGAC 20887

QY      721 AGGTACGGGCGGCTGTGCGCGGCTGGGCGCGCGCGGTCTGCAATTATGAAATATGCC 780
DB      20888 AGGTACGGGCGGCTGTGCGCGGCTGGGCGCGCGCGGTCTGCAATTATGAAATATGCC 20947

QY      781 CACGACATGGGCTATGCGATTCTGACGCGCGGAAACCGGTCCGCTGCA 828
DB      20948 CACGACATGGGCTATGCGATTCTGACGCGCGGAAACCGGTCCGCTGCA 20995

RESULT 5
ABV7558/c
ID      ABV7558 standard; DNA: 36538 BP.
XX      AC      ABV7558;
XX      AC      ABV7558;
XX      DT      22-JAN-2003 (first entry)
XX      DB      Saccharopolyspora butenyl-spinosyn biosynthetic gene cluster 2.
XX      KW      Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrolide;
XX      metabolite; spinosyn; gene; ds.

```

XX		Saccharopolyspora	sp.
OS	Key	Location/Qualifiers	
XX		complement(114. .938)	
FT	CDS	/tag= a	
FT		/product= "busF"	
FT		/note= "No start codon given"	
FT	CDS	1389. .2561	
FT		/tag= b	
FT		/product= "busG"	
FT	CDS	2601. .3353	
FT		/tag= c	
FT		/product= "busH"	
FT	CDS	complement(3359. .4546)	
FT		/tag= d	
FT		/product= "busI"	
FT	CDS	4684. .6303	
FT		/tag= e	
FT		/product= "busJ"	
FT	CDS	6317. .7510	
FT		/tag= f	
FT		/product= "busK"	
FT	CDS	7555. .8406	
FT		/tag= g	
FT		/product= "busL"	
FT	CDS	/note= "No start codon given"	
FT		8640. .9572	
FT		/tag= h	
FT		/product= "busM"	
FT	CDS	/note= "No start codon given"	
FT		complement(9668. .10666)	
FT		/tag= i	
FT		/product= "busN"	
FT	CDS	complement(10675. .12135)	
FT		/tag= j	
FT		/product= "busO"	
FT	CDS	complement(12864. .14177)	
FT		/tag= k	
FT		/product= "busP"	
FT	CDS	14627. .15970	
FT		/tag= l	
FT		/product= "busQ"	
FT		/note= "No start codon given"	
FT	CDS	16008. .17144	
FT		/tag= m	
FT		/product= "busR"	
FT		/note= "No start codon given"	
FT	CDS	17168. .17917	
FT		/tag= n	
FT		/product= "busS"	
FT	CDS	complement(18520. .19932)	
FT		/tag= o	
FT		/product= "ORF LI"	
FT	CDS	complement(19978. .20488)	
FT		/tag= p	
FT		/product= "ORF LI1"	
FT	CDS	/note= "No start codon given"	
FT		complement(20536. .21033)	
FT		/tag= q	
FT		/product= "ORF LI11"	
FT	CDS	/note= "No start codon given"	
FT		21179. .21925	
FT		/tag= r	
FT		/product= "ORF LI1V"	
FT	CDS	complement(22671. .23453)	
FT		/tag= s	
FT		/product= "ORF LI1V1"	
FT	CDS	complement(23687. .24886)	
FT		/tag= t	
FT		/product= "ORF LV11"	
FT	CDS	complement(26177. .26923)	
FT		/tag= u	

FT	/product= "ORF LVIII"
FT	/note= "No start codon given"
FT	27646..28476
FT	/*tag= v
FT	/product= "ORF LIX"
FT	/note= "No start codon given"
XX	
PN	MO200279477-A2.
PD	
PD	10-OCT-2002.
XX	
PF	28-MAR-2002; 2002WO-US009968.
XX	
PR	30-MAR-2001; 2001US-0280175P.
PA	(DOMC ) DOM AGROSCIENCES LLC.
XX	
PI	Hahn DR, Jackson JD, Bullard BS, Gustafson GD, Waldron C;
PI	Mitchell JC;
XX	
DR	WPI: 2003--058434/05.
DR	P-FEDB, ABP57683, ABP57684, ABP57685, ABP57686, ABP57687, ABP57688,
DR	ABP57689, ABP57690, ABP57691, ABP57692, ABP57693, ABP57694, ABP57695,
DR	ABP57696, ABP57697, ABP57698, ABP57699, ABP57700, ABP57701, ABP57702,
DR	ABP57703, ABP57704.
XX	
PT	New butenyl-spirosyn biosynthetic genes, useful for increasing the
PT	production of butenyl-spirosyn insecticidal macroicides, or for changing
PT	the metabolites or products produced by spirosyn-producing
PT	microorganisms.
PS	
XX	Claim 2; Page 99-119; 218pp; English.
XX	
CC	The invention relates to a novel DNA molecule comprising a DNA sequence
CC	that encodes a butenyl-spirosyn biosynthetic enzyme, a butenyl-spirosyn
CC	PKS (polyketide synthetase) domain, or a spirosyn PKS module. The butenyl-
CC	-spirosyn biosynthetic genes are useful for increasing the production of
CC	butenyl-spirosyn insecticidal macroicides. The genes are also useful for
CC	changing the metabolites or products produced by spirosyn-producing
CC	microorganisms. The present sequence represents a DNA molecule encoding
CC	butenyl-spirosyn biosynthetic enzymes
XX	
SQ	Sequence 36538 BP; 6867 A; 12266 C; 11182 G; 6223 T; 0 U; 0 Other;
	Query Match 90.5%; Score 749.6; DB 10; Length 36538;
	Beet Local Similarity 94.1%; Pred. No. 4.8e-15;
	Matches 779; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
OY	
Dd	1 GTGTTGCCAGTGGCGCACCAACATCCGACAGAGTGTTGGGCAGATTAGACTGTGCACG 60
938	GTCGTTGCCAGTGGCGCACCAACATCCGACAGAGTGTTGGGCAGATTAGACTGTGCACC 879
OY	
Dd	61 CCGTTTGCGAAGCTGGTGCAGGAGGCGCCCTCGCCCATTCACCAACGGCTACTGGGAAGAAC 120
878	CCGTTTGCGAAGTTCGTTGCGGAGGCGCCCTCGCCCATTCACCAACGGCTACTGGGAAGAAC 819
OY	
Dd	121 GACGGGCGGGCTTCTCTGGCAGACAGCGCGCAGCGGCTCACCGACTTTGTGCGGAACGG 180
818	GACGGGCGGCAATTCCTTGGCAGACAGCGCGCAGCGGCTCACCGACTTTGTGCGGAACGG 759
OY	
Dd	181 ACCGNGCTCGATGGAGGGGCTTCGACTGCTCATGTGGGGGTGCGGTAACGGACAACAGC 240
758	ACCGGCTCGATGGAGGGGCAATCGACTGCTCATGTGGGGGTGCGGCAACGGACAACAGC 699
OY	
Dd	241 CTGGCGCTGCGCGCGACAACAGCGCATCCAGATCAACGGCATCACGTTGACGACAGTGC 300
698	CTGGCGCTGCGCGCGACAACAGCGCATCCGATCAACGGCATCAACGTTGACGACAGTGC 639
OY	
Dd	301 GTGGCCATCGCGCGCTGATTGGGCAACGGGAACGGGACTTAAGCCAACGGGTGGAATTCTCG 360
638	GCTGCATCGCGCTGATTGGGCGCCCGGAACGGGACTTAAGCCAACAGTGAACCTTCTCG 579
OY	
361	TGCTTCGATGCATATGCTCCCTGCCTACCGGACATGCTTTCGACGCGCGCTGGGCGCATG 420



Db 578 TCCGTCGACGCCATGCTCCCTCCGCTACCCGACATGCTTGGACGCTCTGGGCGCATC 519  
Qy 421 CAGTCGCTGTGAGATGTCGACCGGACCGGATCCGCGGAAATCCTTGCAGTACTC 480  
Db 518 CAGTCGCTGTGAGATGTCGACCGGACCGGATCCGCGGAAATCCTTGCAGTACTC 459  
Qy 481 AAACCCGGTGGCATCTCCGCGCTCACCGAGTGTCTTAAACGAGAACGGGCGCGGAGT 540  
Db 458 AAACCCGGTGGCATCTCCGAGTCAACCGAGTGTCTTAAACGAGAAATAGCGACGGGATC 399  
Qy 541 CCGGTGTCCGGGAGACAGTGGGCGGACCGGCTTCGGATCTGCTGGCTGAGCAATCTTG 600  
Db 398 CCGGTGTCCGGGAGACAGTGGGCGGACCGGCTTCGGATCTGCTGGCTGAGCAATCTTG 339  
Qy 601 GAATCGCTGCTGACGAGCGGCGGATCTGAGATCTGATGGAGAGACGTGTCTGAGAGACC 660  
Db 338 GAATCGCTGCTGAGCGGCGGATCTGAGATCTGATGGAGAGACGTGTCTGAGAGACC 279  
Qy 661 CCGTACTTATGCGCGGATCTCCGCGGAGAGCTGCTGCGCACGAGCGGATCGCGGAC 720  
Db 278 CCGTACTTATGCGCGGATCTCCGCGGAGAGCTGCTGCGCACGAGCGGATCGCGGAA 219  
Qy 721 AGGTACGGGCGCGGCTGCTCCGCGGCGGCGGCGGCTGCGGATTAAGAAATATGCG 780  
Db 218 AGGTACGGGCGCGGCTGCTCCGCGGAGAGCTGCGGCGGCGGCTGCGGATTAAGAAATATGCG 159  
Qy 781 CACGACATGGGCTATGCGATTTCTGACGCGGCGGAGACCGGCTGCGGCTGA 828  
Db 158 GACGACATGGGCTATGCGATTTCTGACGCGGCGGAGACCGGCTGCGGCTGA 111

RESULT 6  
AAD54227  
ID AAD54227 standard; DNA; 846 BP.

XX AAD54227;

DT 17-JUN-2003 (first entry)

DE Streptomyces platensis subspecies rosaceus dorrigocin ORF10 DNA.

KM Polyketide biosynthesis; dorrigocin; DORR; lactimidomycin; LACT; gene;  
da.

OS Streptomyces platensis.

XX Key Location/Qualifiers  
FH 163. .1149  
FT CDS /tag= a  
FT CDS /product= "ORF10 protein"

XX MO200288176-A2.

XX 07-NOV-2002.

XX 26-APR-2002; 2002MO-CA000591.

XX 26-APR-2001; 2001US-0286346P.

XX (ECOP-) ECOPIA BIOSCIENCES INC.

XX Farnet CM, Zazopoulos E, Staffa A, Yang X;

XX MPI; 2003-201222/19.

XX P-PSDB; AAB35493.

PT Novel isolated or purified polypeptide involved in biosynthesis of  
PT polyketide dorrigocin or polyketide lactimidomycin, useful for preparing  
PT dorrigocin or lactimidomycin.

XX Claim 6; Page 198; 312pp; English.

CC The invention relates to novel proteins involved in the biosynthesis of  
CC polyketide dorrigocin (DORR) or lactimidomycin (LACT) biosynthesis by  
CC microorganisms. Sequences of the invention allow direct manipulation of  
CC dorrigocin, lactimidomycin and related chemical structures via chemical  
CC engineering of the enzymes involved in the biosynthesis of dorrigocin and  
CC lactimidomycin. They are useful for introducing chemical handles into  
CC normally inert positions that permit subsequent chemical modifications  
CC and facilitate the development of polyketides. The genes and proteins of  
CC the invention can also be used to generate a focused library of analogues  
CC around a polyketide lead candidate to fine-tune the compound for optimal  
CC properties. They are useful for generating antibodies specific for the  
CC polyketide biosynthesis. The present sequence is Streptomyces platensis  
CC subspecies rosaceus DORR ORF10 DNA

XX SQ Sequence 846 BP; 141 A; 343 C; 249 G; 113 T; 0 U; 0 Other;

Query Match 20.5%; Score 170; DB 10; Length 846;  
Best Local Similarity 59.4%; Pred. No. 5e-27;  
Matches 328; Conservative 0; Mismatches 215; Indels 9; Gaps 2;

Qy 5 TCCGAGTGGGCGGACCAACATGCGAGGTTGGGCAATGATAGCTGTGACGGCGT 64  
Db 35 TCCCGGCTCCGCGCCCGCCGTCGAGAGGTGACACCTTAACGCGCTACCGCAC 94  
Qy 65 TCGTGAATCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 124  
Db 95 TCGACACCGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 148  
Qy 125 GCGGCGGCTTCTGAGC---AGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 181  
Db 149 ACAAGACACCGCGGCTGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 208  
Qy 182 CCGTGTGATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 241  
Db 209 TCGGATTCGACAGGAGACGCGGCTGCGGAGTGGCGGCGGCGGCGGCGGCGGCGG 268  
Qy 242 TCGGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 301  
Db 269 TCGGATTCGACAGGAGACGCGGCTGCGGAGTGGCGGCGGCGGCGGCGGCGGCGG 328  
Qy 302 TGGCATTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 361  
Db 329 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 388  
Qy 362 GCGTGTGATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 421  
Db 389 ACGCCGACCGCATGAACTGCCCTTCCCGAGACTCTTTCGACGCGGCGGCGGCGGCGG 448  
Qy 422 AGTGGCTGTGAGATGTCGGAACCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGG 481  
Db 449 AGTGGCTGTGAGATGTCGGAACCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGG 508  
Qy 482 AAACCCGGTGGCATCTCCGCGCTCACCGAGTGTCTTAAACGAGAACGGGCGCGGAGT 541  
Db 509 GCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 568  
Qy 542 CCGTGTCCGGGCGG 553  
Db 569 AGCAGCCCGCGG 580

RESULT 7

AAD54217  
ID AAD54217 standard; DNA; 52101 BP.

XX AAD54217;

DT 17-JUN-2003 (first entry)

DE Streptomyces platensis subspecies rosaceus dorrigocin DNA.

KM Polyketide biosynthesis; dorrigocin; DORR; lactimidomycin; LACT; gene;  
da.

XX Streptomyces platensis.  
OS  
XX  
PH Key Location/Qualifiers  
FT CDS complement(67..3720)  
FT /tag= a  
FT /product= "ORF1 protein"  
FT CDS 4092..5681  
FT /tag= b  
FT /product= "ORF2 protein"  
FT /note= "No start codon"  
FT /partial  
FT CDS 5767..6018  
FT /tag= c  
FT /product= "ORF3 protein"  
FT 6023..7993  
FT /tag= d  
FT /product= "ORF4 protein"  
FT 8009..17587  
FT /tag= e  
FT /product= "ORF5 protein"  
FT 17634..41714  
FT /tag= f  
FT /product= "ORF6 protein"  
FT /note= "No start codon"  
FT /partial  
FT CDS 41772..47633  
FT /tag= g  
FT /product= "ORF7 protein"  
FT /note= "No start codon"  
FT /partial  
FT CDS 47635..49890  
FT /tag= h  
FT /product= "ORF8 protein"  
FT 49922..50938  
FT /tag= i  
FT /product= "ORF9 protein"  
FT 51234..52079  
FT /tag= j  
FT /product= "ORF10 protein"  
XX  
XX MO200288176-A2.  
XX  
XX 07-NOV-2002.  
XX  
XX 26-APR-2002; 2002MO-CA000591.  
XX  
XX 26-APR-2001; 2001US-0286346P.  
XX  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
XX Farnet CM, Zazopoulos E, Staffa A, Yang X;  
XX  
XX MPI, 2003-201222/19.  
XX P-PDB; AAE35485, AAE35486, AAE35487, AAE35488, AAE35489,  
XX AAE35490, AAE35491, AAE35492, AAE35493.  
XX  
XX Novel isolated or purified polypeptide involved in biosynthesis of  
XX polyketide dorrigocin or polyketide lactimidomycin, useful for preparing  
XX dorrigocin or lactimidomycin.  
XX  
XX Claim 1; Page 85-113; 312pp; English.  
XX  
XX The invention relates to novel proteins involved in the biosynthesis of  
XX polyketide dorrigocin (DORR) or lactimidomycin (LACT) biosynthesis by  
XX microorganisms. Sequences of the invention allow direct manipulation of  
XX dorrigocin, lactimidomycin and related chemical structures via chemical  
XX engineering of the enzymes involved in the biosynthesis of dorrigocin and  
XX lactimidomycin. They are useful for introducing chemical modifications  
XX normally inert positions that permit subsequence chemical modifications  
XX and facilitate the development of polyketides. The genes and proteins of  
XX the invention can also be used to generate a focused library of analogues  
XX around a polyketide lead candidate to fine-tune the compound for optimal

CC properties. They are useful for generating antibodies specific for the  
CC polyketide biosynthesis. The present sequence is Streptomyces platensis  
CC subspecies rosaceus DORR DNA  
XX  
SQ Sequence 52101 BP; 7364 A; 20113 C; 17894 G; 6730 T; 0 U; 0 Other;  
Query Match 20.5%; Score 170; DB 10; Length 52101;  
Best Local Similarity 59.4%; Pred. No. 6.3e-27;  
Matches 328; Conservative 0; Mismatches 215; Indels 9; Gaps 2;  
QY 5 TGGCAGGTGGCGGCACCAACATCGACGAGCTTTGGCCAGATGTATGACTGTGACGCCGT 64  
DB 51268 TCCTCCGTTCGCCCCCTCCCGAGAGTGGACACTTACGACCGCTCACCGCAC 51327  
QY 65 TGCTGACTCGATCGATCGCGGCGGCGCCCTCGGCATTCACCAAGCTACTGTGGAGAACGACG 124  
DB 51328 TGGACATCGGAAGCGGCGCGCG-----GCACTTCACCTTGGCTACTGGAGACGTGACG 51381  
QY 125 GAGCGGCTTCTCTGGC--AGCAGGCGCGCGACCGGCTCACCGACTTGTGCGCGAACGGA 181  
DB 51382 ACAACGACACCCCGCTGTGGAGAGGCGCGACCGGCTCACCGACGATGACCGACCGCC 51441  
QY 182 CCGTCTGATGAGCGGCGCTTGCATGCTCTCATGTGGGTGGGTACCGGACCAACGACGCGC 241  
DB 51442 TGGCGATGACCAAGGACAGGAGCTCTCGACGTGGGTGGAGTGGCGGACCGGCGCA 51501  
QY 242 TGCGGCTGCGCGCGACCAACGCGATCCAGATCACCGGATCACCGTCCAGCGGATGACG 301  
DB 51502 TGGCGATGCGCGCGCGACCGCGCGCCCATGTGCACGGGATGCCATACGAAAGACCGA 51561  
QY 302 TGGCCATGCGCGCTGATTTGCGCACGCGAACGCGGACTAAGCACCGGATGAGCTTCTGT 361  
DB 51562 TCGCCCGCGCACCGCCCTCGCCGAGAGGCGCGGCGCTGAGGACCGCGTGAATTCGCGC 51621  
QY 362 GCGTGCATGATGATGCTCCCTGCGCGTACCGCGACATGCTTTGAGACCGCGCTGGCCATGCG 421  
DB 51622 ACGCGACGCGATGAACTGCTCTTCCCGACGACTCTTTCGACGCGCGCATATCGCATCG 51681  
QY 422 AGTCGCTGTGAGAGATGTCGGAACCGGACCGTGCATTCGGGAATCTTTCGATGACTCA 481  
DB 51682 AGTCGATTTCCACATGCGCGACCGCGGACGGGTCTCTCGCGAATTCGCGCGCTACTCGC 51741  
QY 482 AACCGGTGCGATCTCTGCGCGTCAACGAGTGTCTAAACGAAAGCGGCGCGGAGATGC 541  
DB 51742 GCCCGGCGCGCGCTGTGCTCTTCCACGACTTCTTCGAGCGCGCGCTCCCGCCGAGA 51801  
QY 542 CGGTGTCCGGGG 553  
DB 51802 AGCAGCCCGCG 51813  
RESULT 8  
AAFB8324  
ID AAFB8324 standard; DNA; 900 BP.  
XX  
XX AAFB8324;  
XX  
XX 28-AUG-2001 (first entry)  
XX  
XX S. spinosa DNA fragment encoding ORF6, SEQ ID 19.  
XX  
XX DE  
XX  
XX Forosamine; trimethylxanthose; polyketide synthase; biosynthesis;  
XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;  
XX macrolide; insecticidal; O-methyltransferase; ds.  
XX  
XX Saccharopolyspora spinosa.  
XX  
XX DE19957268-A1.  
XX  
XX 08-MAR-2001.  
XX  
XX 29-NOV-1999; 99DF-01057268.  
XX

PR 27-AUG-1999; 99DE-01040596.  
 XX (FARB ) BAYER AG.  
 PA  
 XX  
 PI Eberz G, Moehle V, Froede R, Velten R, Salas JA;  
 XX WPI, 2001-267102/28.  
 DR P-PSDB; AAB70954.  
 XX  
 PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for  
 PT recombinant production of insecticidal spinosyns and their derivatives.  
 XX  
 PS Claim 7a; Page 126-127; 354pp; German.  
 XX  
 XX This invention describes a novel method nucleic acid (I) and its encoded  
 CC polypeptide (II) containing at least one region that encodes an enzymatic  
 CC activity involved in biosynthesis of spinosyns. (I) are used (i) to  
 CC identify, inactivate or modulate genes involved in the biosynthesis of  
 CC (II); (ii) to generate a library of polyketide synthases; (iii) for  
 CC adding forosamine or trimethylamine to a spinosyn or polyketide  
 CC aglycone; and (iv) for recombinant production of the corresponding  
 CC derivatives, including production of transgenic plants that express (II)  
 CC and thus have increased resistance to insects. (I) are also useful as  
 CC markers for sequencing of the *Saccharopolyspora spinosa* genome. (II) are  
 CC macrolides with insecticidal, but not antibacterial, activity, and can  
 CC also be used to raise specific antibodies, useful for identifying  
 CC expression clones in a gene bank. Cells transformed with (I) may produce  
 CC (II) at significantly increased levels or produce new derivatives of  
 CC (II). This sequence, ORF 6, encodes an S. spinosa O-methyltransferase  
 CC protein  
 XX  
 SQ Sequence 900 BP; 172 A; 292 C; 289 G; 147 T; 0 U; 0 Other;  
 Query Match 20.0%; Score 165.4; DB 4; Length 900;  
 Best Local Similarity 52.8%; Pred. No. 4.8e-26;  
 Matches 386; Conservative 0; Mismatches 336; Indels 9; Gaps 1;

QY 15 CGCACCACATCCGACAGGTTGGGAGATGATGACCTGTGACCGCGTTGCTGAATC 74  
 DB 21 CACACAGGCAATTCGACCGCGATGAGTGATCATCTTCGATCGTTGGCGACGG 80  
 QY 75 GGTGCGGGGCGGCGCCCTGCGCCATCCACACCGGCTACTGGAGAGACGCGGGGCTTC 134  
 DB 81 GCGTCCCTTCACACCGGTTACTGGGCGCGGATACGGAGAGATGCCGGTCCACACC 140  
 QY 135 CTGGACGACGAGCGCGGCGGACCGGCTCACTTGTGCGGAAAGGACCGGTCTCATG 194  
 DB 141 GTGATCGATGCTGCGACCACTGACCGACTGTTTATGACAAAGCCCGGCTCC 200  
 QY 195 CGGCGTTGCACTGCTCATGTGGGGTGGGTAACGCAACAGCGCTGCGCGTCCGCG 254  
 DB 201 CGAGAGGCACTGTTGCACTGGGGCTGGCAATGGGCAAGCCGTAAGTCCGTGGGATG 260  
 QY 255 GCAACAACGCATCCAGATCACCGGCACTACCGTACGACAGTGAAGTGGCCATCGCGC 314  
 DB 261 CGGCAAGGCGGCTTGAATCACCGGAATCACCGTGAAGCCGACGATCTCCCGCGCAC 320  
 QY 315 TGATTGGGCAAGGCAAGCGGACTAAGCAACCGGGTGAATCTTCGTGCGTCAATGCGCAT 374  
 DB 321 CAGGCTGCGCAACGAGACCGGACTGGCGGCGGCTTGAAGTTGATTAATTCAGCGCGC 380  
 QY 375 GTCCCTGCGTACCGGACATGCTTTCAGACGCGGCTGGGCGCATGCAATGCGTGTGA 434  
 DB 381 CCAAGTCCCTTACCGGACGCGTTTCTTTCAGGCGCGGATGGCGATGCAATGCGTGTGA 440  
 QY 435 GATGTCCGAACCGGACCGTGCATCCGGAATCTTTCAGTACTCAAAACCGGTGGCAT 494  
 DB 441 GATGTGGAACCGGACCGGCGGATCCGAGATCCGAGATCTTGGAACCGGCGGCG 500  
 QY 495 CCGTGGGCTACCGGAGTGTCTCAACGAGAGCGGGCGGCGGATGCGGATGTCGCGGGA 554  
 DB 501 GTTGTCTCTGGAGACATCACTCTCGGTTTCGACTCCGGAAGATACCGCGCGGTTTG 560

QY 555 CAGTGGCCGACCGGCGCTTCGATCTGCTGCTGAGCACTTGTGAATGCTGCTGTC 614  
 DB 561 GACGGGACACCGGCGCCATACCTT-----GAAACGCTTACAGCGCGGTGTACGGA 611  
 QY 615 ACCGGGGTTGAGATCTCGATTTGGAGAGACGTGTGTCGAGAACCCGGTACTTATGCC 674  
 DB 612 AGCCGGGTTGAGATTTCTGATGATGACGACCTCAAGGACAGACAGGTGATGTCTC 671  
 QY 675 GCAATTGCGGAAGAGCTGCTGCGCACGACGAGATCGCGACAGGTACGGGCGGCG 734  
 DB 672 CTGTACGTCGACGAGTTGCTCCGAAACTGATGAGTCTGCGGCGTTCGACCTCCGCG 731  
 QY 735 TGTCCCGCGCT 745  
 DB 732 TGTCCGCACT 742

RESULT 9  
 AAT70153  
 ID AAT70153 standard; DNA; 6085 BP.  
 AC AAT70153;  
 XX  
 DT 17-OCT-2003 (revised)  
 DT 23-OUT-1997 (first entry)  
 XX  
 DE S.longisporoflavus staurosporin synthesis gene cluster 6.5kb fragment.  
 KW Staurosporin; gene cluster; indole-carbazole alkaloid; ICA; antibiotic;  
 KW antiproliferative; platelet aggregation; fungus; yeast;  
 KW Ca2+/phospholipid-dependent serine/threonine protein kinase; ds.  
 XX  
 OS Streptomyces longisporoflavus; (strain R19).  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 378..1655  
 FT /tag= a  
 FT /label= Gene 1  
 FT /note= "Encodes a protein containing 425 amino acids"  
 FT 1747..2553  
 FT /tag= b  
 FT /label= Gene 2  
 FT /note= "Encodes a protein containing 268 amino acids  
 FT significantly similar to known S-adenosyl methionine-  
 FT dependent methyl transferase"  
 FT 2593..4011  
 FT /tag= c  
 FT /label= Gene 3  
 FT /note= "Encodes a protein containing 472 amino acids"  
 FT 4013..4959  
 FT /tag= d  
 FT /label= Gene 4  
 FT /note= "Encodes a protein containing 328 amino acids"  
 FT 5071..6085  
 FT /tag= e  
 FT /label= Gene 5 (part)  
 FT /note= "Partial sequence for gene 5, full length protein  
 FT contains 366 amino acids and is significantly similar to  
 FT amino transferase enzymes, such as the Dnr J protein"  
 PN MO9708323-A1.  
 XX  
 XX 06-MAR-1997.  
 PD  
 XX  
 PF 19-AUG-1996; 96MO-BP003643.  
 XX  
 PR 30-AUG-1995; 95EP-00810534.  
 XX  
 XX (CIBA ) CIBA GEIGY AG.  
 PA  
 PI Schupp T, Engel N, Bietenhader J, Toupet C, Pospishech A;

DR WPI, 1997-179280/16.  
XX Indole-carbazole alkaloid biosynthesis gene cluster - especially coding  
PT for the antibiotic staurosporin from Streptomyces longisporolavus.  
XX  
PS Claim 7, Page 37-41; 55pp; English.  
XX  
XX The present sequence represents the 6.5kb PvuII fragment of Streptomyces  
CC longisporolavus R19 which is involved in the biosynthesis of indole-  
CC carbazole alkaloids (ICA). The sequence contains five functional  
CC fragments as indicated in the features table. The DNA or a hybrid vector  
CC containing it can be used to prepare an ICA or derivatives and  
CC precursors, either by allowing production in previously incapable  
CC organisms or by improving yields. In particular, the antibiotic  
CC staurosporin can be produced. Staurosporin is known to have inhibitory  
CC activity against fungi, yeasts, and G2+/phospholipid-dependent  
CC serine/threonine protein kinases (PKCs). Staurosporin also has  
CC antiproliferative activity and can inhibit platelet aggregation. The  
CC present sequence can also inactivate ICA biosynthesis genes and can be  
CC used in PCR amplification. An advantage of this is that productivity of  
CC staurosporin-synthesizing Streptomyces is improved over natural strains  
CC yielding only low concentrations. (Updated on 17-OCT-2003 to standardise  
XX OS field)

XX Sequence 6085 BP, 882 A; 2374 C; 1992 G; 833 T; 0 U; 4 Other;

XX Query Match 20.0%; Score 165.4; DB 2; Length 6085;

XX Best Local Similarity 60.6%; Pred. No. 5.4e-26;

XX Matches 271; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

OY 90 CTGCGCCATCCACACGCGCTACTGTGGAGAACGACGCGGGCTTCTTGGACAGAGCGCGC 149  
DB 1833 CGGCAACATCCACGCGGGTACTGGAGACGACCCCGAGGTCGATGCGCGAGGCAC 1892  
OY 150 CGACCGGCTCACCGGACCTTTCGCGGAAACGACGCTGTCATGGCGCGCTTCCATCTCT 209  
DB 1893 CGACCGGCTCACCGGACCTTTCGCGGAAACGACGCTGTCATGGCGCGCTTCCATCTCT 1952  
OY 210 CGATGTGGGGTTCGCGTACCGGACACGACGCTGCGCGCGCGACGACGCGATCTCA 269  
DB 1993 GGAAGTGGGGTTCGCGTACCGGACACGACGCTGCGCGCGCGACGACGCGATCTCA 2012  
OY 270 GATCAACCGGATCACCGTCAAGCAGGTGCAAGTGGCCATTCGCGCTGATTGGCAGCGCA 329  
DB 2013 CGTCAACCGGATCACCGTCAAGCAGGTGCAAGTGGCCATTCGCGCTGATTGGCAGCGCA 2072  
OY 330 ACGCGGATCAAGCAGCGGGTGAATCTTCGCGGTGATGCAATGCTCCGCTGACCC 389  
DB 2073 GTCCGATCCGCGGGGCGGGTCTCTTCGCGGTGCGGACGCGCATGACCTTCTTCA 2132  
OY 390 GGACATGCTTTCGACGCGCGCTGGGCGCATGCTGCTTGGAGATGTCGACCGGA 449  
DB 2133 GGAAGTCTCTTCGACGCGGGCTTTCGCGGTGCGGACGCGCATGCTGCTTGGAGATGTCGACCGGA 2192  
OY 450 CCGTCCATTCGCGGAAATCTTTCGAGTACTCAACCCGGTGGCATCTTCGCGGTCAACCGA 509  
DB 2193 ACCCGCGTCAAGGATGACCGGGGTGCTCGCGCGCGCGCGCTTCATGATGCGCGA 2252  
OY 510 GGTGCTCAAGAGAGGCGGGCGCGG 536  
DB 2253 CCGTGTGACGACAGCGGTTTACCGG 2279

RESULT 10

AAF88315 standard; DNA, 45624 BP.

AAF88315;

28-AUG-2001 (first entry)

S. epinosa DNA fragment SEQ ID 4.

KW Forosamine; trimethylrhinamose; polyketide synthase; biosynthesis;  
XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;  
KW macrolide; insecticidal; ds.  
XX  
OS Saccharopolyspora epinosa.  
XX  
XX DE19957268-A1.  
XX  
XX 08-MAR-2001.  
XX  
XX 29-NOV-1999; 99DE-01057268.  
XX  
XX 27-AUG-1999; 99DE-01040596.  
XX  
XX (FARB ) BAYER AG.  
XX  
PI Ebertz G, Moehrle V, Froede R, Velten R, Salas JA;  
XX  
DR WPI, 2001-267102/28.  
XX  
XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for  
XX recombinant production of insecticidal spinosyns and their derivatives.

XX Claim 7, Page 58-74; 354pp; German.

XX This invention describes a novel method nucleic acid (I) and its encoded  
CC polypeptide (II) containing at least one region that encodes an enzymatic  
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to  
CC identify, inactivate or modulate genes involved in the biosynthesis of  
CC (II); (ii) to generate a library of polyketide synthases; (iii) for  
CC adding forosamine or trimethylrhinamose to a spinosyn or polyketide  
CC aglycone; and (iv) for recombinant production of the corresponding  
CC enzymes, which are used for production of (II), their precursors or  
CC derivatives, including production of transgenic plants that express (II)  
CC and thus have increased resistance to insects. (I) are also useful as  
CC markers for sequencing of the Saccharopolyspora epinosa genome. (II) are  
CC macrocyclic with insecticidal, but not antibacterial, activity, and can  
CC also be used to raise specific antibodies, useful for identifying  
CC expression clones in a gene bank. Cells transformed with (I) may produce  
CC (II) at significantly increased levels or produce new derivatives of  
CC (II). This sequence represents a genomic DNA fragment of the S. epinosa  
CC genome which contains the coding regions for proteins involved in  
CC forosamine and trimethylrhinamose biosynthesis

XX Sequence 45624 BP, 7933 A; 14913 C; 14940 G; 7838 T; 0 U; 0 Other;

XX Query Match 20.0%; Score 165.4; DB 4; Length 45624;

XX Best Local Similarity 52.8%; Pred. No. 6.1e-26;

XX Matches 386; Conservative 0; Mismatches 336; Indels 9; Gaps 1;

OY 15 CGACCAACATCGACGAGGTGGGACAGATGATGACCTGGTCAAGCGCTTGTGTAATCTC 74  
DB 7421 CACACAGGCAATTTCCGACCGCGGATCAGTGAGATCATCTTCGATGCGGTGGCGACGG 7480  
OY 75 GGTGCGGGGCGGCGCTTCGCGCATTCACACGCTACTGGAGAACGACGCGGGCTTC 134  
DB 7481 GGTGCGGGGCGGCGCTTCGCGCATTCACACGCTACTGGAGAACGACGCGGGCTTC 7540  
OY 135 CTGGACACAGGCGCGGACCGGCTCAACGACCTTTCGCGGAAACGACGCGGTCCGATGG 194  
DB 7541 GTGGTCCGATGCTGCGACCACTGATGACGACCTGTTATGACCAAGGCGCGGCTCGCTC 7600  
OY 195 CGGCTTCACTGCTCGATGATGGGGTTCGCGTACCGGACCAACGACGCTGCGGTGCGCG 254  
DB 7601 CGGAGCGCACTGTTTCACTGCGGCTCGCGCAATGGGCAAGCCGTATCCGTGGCGCATG 7660  
OY 255 CGACACGCGATCCAGTCAACCGGATCACCGGCTGACGCAAGTGCAGTGGCCATCGCGC 314  
DB 7661 CGCCAGCGGGCTTTCGAGTCAACCGGAATCACCGTGAAGCCGACGATCTCGCGCGCAC 7720  
OY 315 TGATTGGCAGCGGAGCGGACGTAAGCAACCGGATGATCTTCGCGCGGTGCGATGCGCAT 374  
DB 7721 CAGGCTCGCAACGACGACGCGGACTGGCGGCACTTTCGAGTTGATCTGATGCGCGCGC 7780



[illegible]

```

FT misc_feature 37108..38097
FT /tag= ah
FT /note= "Acyl transferase domain (AT3) : part of extender
FT module 3"
FT CDS 38916..35374
FT /tag= y
FT /product= "gpnC"
FT /note= "Spinosyn biosynthesis protein C"
FT 38992..39528
FT /tag= a1
FT /note= "Beta-ketoreductase domain (KR3) : part of extender
FT module 3"
FT 39790..40035
FT /tag= aJ
FT /note= "Acyl carrier protein domain (ACP3) : part of
FT extender module 3"
FT 40102..41373
FT /tag= ak
FT /note= "Beta-ketosynthase domain (KS4) : part of extender
FT module 4"
FT 41713..42705
FT /tag= a1
FT /note= "Acyl transferase domain (AT4) : part of extender
FT module 4"
FT 43615..44157
FT /tag= am
FT /note= "Beta-ketoreductase domain (KR4) : part of extender
FT module 4"
FT misc_feature 44431..44676
FT /tag= an
FT /note= "Acyl carrier protein domain (ACP4) : part of
FT extender module 4"
FT 44966..459752
FT /tag= ao
FT /product= "apnD"
FT /note= "Spinosyn biosynthesis protein D"
FT 45077..46348
FT /tag= ap
FT /note= "Beta-ketoreductase domain (KS5) : part of extender
FT module 5"
FT 46691..47674
FT /tag= ag
FT /note= "Acyl transferase domain (AT5) : part of extender
FT module 5"
FT misc_feature 47753..48310
FT /tag= ar
FT /note= "Dehydratase domain (DH5) : part of extender module
FT 5"
FT 49226..49771
FT /tag= as
FT /note= "Beta-ketoreductase domain (KR5) : part of extender
FT module 5"
FT misc_feature 50009..50254
FT

Query Match 20.0%; Score 165.4; DB 2; Length 80161;
Best Local Similarity 52.8%; Pred. No. 6.3e-26;
Matches 386; Conservative 0; Mismatches 336; Indels 9; Gaps 1;

QY 15 CGACCAACATCCGACAGGTTGGGACAGATGATGACCTGTACCGCGTTGCTGAATC 74
DB 13575 CACACAGGCAATTCGACCGCGATCAGGTGAGTCCATCTTGATGCGTTGGCGCACGG 13516
QY 75 GGTGCGGGGCGCCCTGCGCATCCACACGAGTACTGGAGAAACGCGGGCGGCTTC 134
DB 13515 GCGTCCCTCGACACAGGTTACTGGGCGGCGGTATCGGAGGATGCCGATGCCACAC 13456
QY 135 CTGGCAGACGAGCGCGGCTCACCGGCTTGTGTGCGCGAAGGACCGTGTCCATGG 134
DB 13455 GTGTGCGATGTGCTGCCACCACTGACCTGTTATGACAAAGGCGCGCTCCGTC 13396
QY 195 CGGGTTGCACTGCTGATGTGGGTGCGGTACCGGACCAACGAGCGTGCCTGCGCGC 254
DB 13395 CGAGCGCACTGTTCACTGGGCTGCGGCAATGGGACGCCCTGATCCGTGCGGATG 13336

```

```

QY 255 CGACCAACGCGATCCGATCACCGGCATCACCGTACGACGAGGTGCAGTGGCCATGCCGC 314
DB 13335 CGCCAGCGCGCTTCGATGATCACCGGAATCACCGTGAAGCCGACGATCTCGCGCGGCAC 13276
QY 315 TGATTCGACGCGGACCGGACGCTAAGCCACCGGTTGACTTCTGTGCTGATGCCAT 374
DB 13275 CAGGCTCCGCAACGAGACCGGACTGCGCGGACGTCTGATGATCTGATCTGACAGGCGC 13216
QY 375 GTCCCTGCGGATCCCGGACAAATGCTTTCGACGCGCGCCCTGGGCGCATGAGTGGCTTGA 434
DB 13215 CCAGCTGCGCTTACCGGACGAGTCTTTCAGGCGCGCATGGGAGATGACGTCGTGTGA 13156
QY 435 GATGTCGACCGGACCGGCGCATCCGGAATCTTGTGAGTACTCAACCGCGTGGCAT 494
DB 13155 GATTCGTGACGAGCGCGCGCGATCCGAGGTCCACGGAATCTGGAACCGGCGCGCG 13096
QY 495 CCTCGGCGTCAACGAGTGTCTAAACGAAACGCGGCGGAGATGCCGTGTCCGGGGA 554
DB 13095 GTTCGTCTCGGAGACATCATCATCTCGGTTTCGATCCCGGAAGATACGCGCGGTTTG 13036
QY 555 CAGTGGCCGACCGGCTTTCGATCTGCTGCTGCTGAGCAACTTCTGGAATGCTGCTGTC 614
DB 13035 GACGGGCGACGACCGGCTTACCTT-----GAACAGCTTCACGCGCGTGTGACGA 12985
QY 615 AGCGGGGTTGAGATCTCGATTTGGAGAGACGTGTGTCGAGGACCGGTAATTGATGCC 674
DB 12984 AGCCGGGTTGAGATTTCTGAGTACAGTACGACCTTCAAGGACGACGACGATGATGCTTC 12925
QY 675 GCAGTTCGCGGAAGTCTGCTGCGCACGACGACGAGATGCGGACAGTACGCGCGCGC 734
DB 12924 CTGGTACGTGACGAGTGTCTCGGAACCTGATGAGTCTCGCGGCTGACGCTGCGCGC 12865
QY 735 TGTGCGCGGCT 745
DB 12864 TGTGCGCACCT 12854

RESULT 13
ABV75558
ID ABV75558 standard; DNA; 36538 BP.
XX
AC ABV75558;
XX
DT 22-JAN-2003 (first entry)
XX
DE Saccharopolyspora butenyl-spinosyn biosynthetic gene cluster 2.
XX
KW Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrolide;
KW metabolite; spinosyn; gene; ds.
OS
XX Saccharopolyspora sp.
XX
XX Key Location/Qualifiers
FT CDS complement(114..938)
FT /tag= a
FT /product= "busF"
FT /note= "No start codon given"
FT 1389..2561
FT /tag= b
FT /product= "busG"
FT 2601..3353
FT /tag= c
FT /product= "busH"
FT /product= "busI"
FT /tag= d
FT complement(3359..4546)
FT /tag= e
FT /product= "busJ"
FT 6317..7510
FT /tag= f
FT /product= "busK"

```

FT	CDS	7555. .8406
FT	/tag= g	
FT	/product= "buel"	
FT	/note= "No start codon given"	
FT	CDS	8640. .9572
FT	/tag= h	
FT	/product= "buam"	
FT	/note= "No start codon given"	
FT	complement(9668. .10666)	
FT	/tag= i	
FT	/product= "buen"	
FT	complement(10675. .12135)	
FT	/tag= j	
FT	/product= "busO"	
FT	complement(12864. .14177)	
FT	/tag= k	
FT	/product= "busP"	
FT	14627. .15970	
FT	/tag= l	
FT	/product= "busQ"	
FT	/note= "No start codon given"	
FT	16008. .17144	
FT	/tag= m	
FT	/product= "busR"	
FT	/note= "No start codon given"	
FT	17168. .17917	
FT	/tag= n	
FT	/product= "busS"	
FT	complement(18520. .19932)	
FT	/tag= o	
FT	/product= "ORF LI"	
FT	complement(19978. .20488)	
FT	/tag= p	
FT	/product= "ORF LII"	
FT	/note= "No start codon given"	
FT	complement(20536. .21033)	
FT	/tag= q	
FT	/product= "ORF LIII"	
FT	/note= "No start codon given"	
FT	21179. .21925	
FT	/tag= r	
FT	/product= "ORF LIV"	
FT	complement(22671. .23453)	
FT	/tag= s	
FT	/product= "ORF LVI"	
FT	complement(23687. .24866)	
FT	/tag= t	
FT	/product= "ORF LVII"	
FT	complement(26177. .26923)	
FT	/tag= u	
FT	/product= "ORF LVIII"	
FT	/note= "No start codon given"	
FT	27646. .28476	
FT	/tag= v	
FT	/product= "ORF LIX"	
FT	/note= "No start codon given"	
PN	W0200279477-A2.	
XX		
PD	10-OCT-2002.	
XX		
PE	28-MAR-2002; 2002WO-US009368.	
XX		
PR	30-MAR-2001; 2001US-0280175P.	
XX		
PA	(DOWC ) DOW AGROSCIENCES LLC.	
PI	Hahn DR, Jackson JD, Bullard BS, Gustafson GD, Waldron C;	
XX	Mitchell JC;	
DR	WPI; 2003-058434/05.	
DR	P-BSDBI; ABP57683, ABP57684, ABP57685, ABP57686, ABP57687, ABP57688,	
DR	ABP57689, ABP57690, ABP57691, ABP57692, ABP57693, ABP57694, ABP57695,	

DR	ABP576566, ABP57667, ABP576598, ABP57659, ABP577001, ABP57702, ABP57703, ABP57704.
XX	New butenyl-spinosyn biosynthetic genes, useful for increasing the
PT	production of butenyl-spinosyn insecticidal macrocides, or for changing
PT	the metabolites or products produced by spinosyn-producing
XX	microorganisms.
XX	
PS	Claim 2; Page 99-119, 218pp; English.
XX	
CC	The invention relates to a novel DNA molecule comprising a DNA sequence
CC	that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn
CC	PKS (polyketide synthetase) domain, or a spinosyn PKS module. The butenyl-
CC	-spinosyn biosynthetic genes are useful for increasing the production of
CC	butenyl-spinosyn insecticidal macrocides. The genes are also useful for
CC	changing the metabolites or products produced by spinosyn-producing
CC	microorganisms. The present sequence represents a DNA molecule encoding
XX	butenyl-spinosyn biosynthetic enzymes
XX	
SQ	Sequence 36538 BP; 6867 A; 12266 C; 11192 G; 6223 T; 0 U; 0 Other;
	Query Match 19.2%; Score 159; DB 10; Length 36538;
	Best Local Similarity 52.3%; Pred. No. 1.4e-24;
	Matches 382; Conservative 0; Mismatches 340; Indels 9; Gaps 1;
QY	15 CGACCAACATCGACGAGGTGGGAGATGTATGACCTGTGTACGCCGTTGCTGAATCTC 74
DB	7527 CGCACAGGCTTTCCAAAGCGCGGATCAAGTGAAGTCATCTTTCGACGCGTTGGCGCAAGG 7586
QY	75 GGTGCGGGGGCGGCGCCCTTCGCGCATTCACCAACGGCTACTGAGGAAACAGCGGGGGCTTC 134
DB	7587 GCGTGCCTTGACACACGATATCGGGCGGGGGGATGGGAGATGCGGGGGCCACAC 7648
QY	135 CTGGAGAGAGGCGCGCGACCGGCTCACCGACCTTGTGCGCGAAGGACCGACGCGTGTGATGG 194
DB	7647 TTGTGCGAAGCTGTGCGACCACTGACCTGATCTGTTATCGAACAGGCGCGCTTCGCGCC 7706
QY	195 CGCGCTTGCATGCTGATGTGTGGGGTGGCTGATCCGACCAACGACGCGTGCCTGCGCG 254
DB	7707 CGAGCGCACCTGTTGCACTGTGGGCTGTGGCAATGGGCAAGCCCGATGTCGCGGCGACG 7766
QY	225 GCACACAGCGATCCAGATCACCGGCATACCGTCAAGCCAGGTGCAAGTGGCCATTCGCCG 314
DB	7767 CACCAAAAGCGTTGAGATCACCGGATCACCGTGAACCCGCAACATCTCGCGCGCTAC 7828
QY	315 TGATTGGCGACGCGGAAGCGGACTAAGCACCGGGTGAATCTTCGTGCTGTGATGGCAT 374
DB	7827 CAGGCTGCCCAACGAGACCGGACCTGGCCGACAGTCTTGTTCGATCTGATGTGACGCGC 7888
QY	375 GTCCCTGCGGTACCCGGAACAATGTCTTTCGACGCGCGCTGGGCGCAATGACGTGCTGTGA 434
DB	7887 CCGGTGCTTACCCGGAAGGTTCTTTCACGCGGCAATGGCGATGCGATCCGATGTA 7946
QY	435 GATGTCGACACCGACCGGTGCAATCCGCGGAAATCTTTCGAGTACTCAAAACCGGTGGCAT 494
DB	7947 GATGTCGACACCGGTGCGGATCCGCGATCCGGAAGTCAACCGAATCTTTCGAAACCGCGCGCA 8006
QY	495 CCTGCGCGTCAACGAGTGTCTCAAAACGAAGCGCGGCGCGGAGTGCCTGTGTCGCGGA 554
DB	8007 GTTGTCTCTGGGGAACATCATCTGCTGTGCACTCCCGGAAGATACGCGGCGTTTG 8066
QY	555 CAGGTGCGCAACCGGCTTTCGATCTGCTGCTGAGCAATTTCTGGAATCGTGTGCTGCG 614
DB	8067 GACCGGACGACCGGCTTACCTT-----GACAGCTTCAACCGGCTGTGTAAGGA 8117
QY	615 AGCGGGTTGAGATCTCTGATTTGGAGAGAGTGTGTGAGAGACCGGTAATCTTATGCGC 674
DB	8118 AGCGGGTTGAGATTTCTGAAATCACCGACCTCACCGGCGCAACAGATGCATGTGTCTC 8177
QY	675 GCAGTTCGCGCAAGAGCTGCTGGCGACACGAGGATCGCGGACAGATACGGGCGCGC 734
DB	8178 CTGTGATGTGACAGAGTTGCTCCGAGAACTTCGATGAGAGTCCCGGCGTGTGAGCGCGC 8237



```

QY      735 TGTGCGGAGCT 745
DB      8238 TGTGCGGAGCT 8248

RESULT 14
ADJ53165
ID      ADJ53165 standard; DNA; 25681 BP.
XX
XX      ADJ53165;
AC
XX
XX      06-MAY-2004 (first entry)
DT
XX
XX      Saccharothrix aerocolonigenes rebeccamycin biosynthetic gene cluster.
DE
XX
XX      Rebeccamycin biosynthetic gene cluster; rebeccamycin production;
KM      indolocarbazole production; ATCC39243; Streptomyces albus;
KM      antitumor agent; cytostatic; Gram positive bacterium; antibacterial;
XX      antibiotic; gene; ds.
XX
XX      Lechevalieria aerocolonigenes; ATCC39243.
OS
XX
XX      Key
FH      Location/Qualifiers
FT      2..136
FT      /*tag= a
FT      /partial
FT      /product= "OrfD13 protein (SEQ ID NO:2)"
FT      /note= "No start codon given"
FT      complement(302..3313)
FT      /*tag= b
FT      /product= "OrfR5 protein (SEQ ID NO:3)"
FT      /function= "Regulatory protein"
FT      /transl_except= (pos:3310..3313,aa:Met)
FT      complement(3395..4027)
FT      /*tag= c
FT      /product= "OrfR4 protein (SEQ ID NO:4)"
FT      /function= "Dipeptidase"
FT      4402..5718
FT      /*tag= d
FT      /product= "OrfD1 protein (SEQ ID NO:5)"
FT      /function= "Sesterase"
FT      complement(5946..6347)
FT      /*tag= e
FT      /product= "OrfR3 protein (SEQ ID NO:6)"
FT      6581..7768
FT      /*tag= f
FT      /product= "OrfD2 protein (SEQ ID NO:7)"
FT      complement(7841..9106)
FT      /*tag= g
FT      /product= "OrfR2/ngt protein (SEQ ID NO:8)"
FT      /function= "N-glucosyltransferase"
FT      9316..10737
FT      /*tag= h
FT      /product= "OrfD3 protein (SEQ ID NO:9)"
FT      /function= "L-tryptophan oxidase"
FT      10734..13775
FT      /*tag= i
FT      /product= "OrfD4 protein (SEQ ID NO:10)"
FT      13772..15361
FT      /*tag= j
FT      /product= "OrfD5 protein (SEQ ID NO:11)"
FT      /function= "Monooxygenase"
FT      15358..16551
FT      /*tag= k
FT      /product= "OrfD6 protein (SEQ ID NO:12)"
FT      /function= "Cytochrome P450"
FT      /transl_except= (pos:15358..15360,aa:Met)
FT      16578..17399
FT      /*tag= l
FT      /product= "OrfD7 protein (SEQ ID NO:13)"
FT      /function= "Methyltransferase"
FT      /transl_except= (pos:16578..16580,aa:Met)
FT      17730..20501
FT      CDS

```

```

FT      /*tag= m
FT      /product= "OrfD8 protein (SEQ ID NO:14)"
FT      /function= "Regulatory protein"
FT      20498..21010
FT      /*tag= n
FT      /product= "OrfD9 protein (SEQ ID NO:15)"
FT      /function= "Flavin reductase"
FT      21007..22287
FT      /*tag= o
FT      /product= "OrfD10 protein (SEQ ID NO:16)"
FT      /function= "Membrane transport protein"
FT      22271..23863
FT      /*tag= p
FT      /product= "OrfD11 protein (SEQ ID NO:17)"
FT      /function= "Tryptophan halogenase"
FT      complement(23933..25354)
FT      /*tag= q
FT      /product= "OrfR1 protein (SEQ ID NO:18)"
FT      /function= "Membrane transport protein"
FT      25439..25681
FT      /*tag= r
FT      /partial
FT      /product= "OrfD12 protein (SEQ ID NO:19)"
FT      /function= "Regulatory protein"
FT      /note= "No stop codon given"
FT      /transl_except= (pos:25439..25441,aa:Met)
XX
XX      MO2003033706-A1.
XX
XX      24-APR-2003.
XX
XX      17-OCT-2002; 2002MO-ES000492.
XX
XX      19-OCT-2001; 2001ES-00002312.
XX
XX      (UYOV-) UNITV OVIEDO.
XX
XX      Sanchez Reillo C, Fernandez Brana A, Salas Fernandez JA;
XX      Mendez Fernandez C;
XX
XX      WPI; 2003-393533/37.
XX      DR      P-PSDB; ADJ53166, ADJ53167, ADJ53168, ADJ53169, ADJ53170, ADJ53171,
XX      DR      ADJ53172, ADJ53173, ADJ53174, ADJ53175, ADJ53176, ADJ53177, ADJ53178,
XX      DR      ADJ53179, ADJ53180, ADJ53181, ADJ53182, ADJ53183.
XX
XX      Production of indolo-carbazole antitumor agents, especially rebeccamycin
XX      or derivatives, by culturing host cells transformed with vector
XX      containing DNA from Saccharothrix aerocolonigenes.
XX
XX      Claim 2; SEQ ID NO 1; 76pp; Spanish.
XX
XX      The invention relates to a method for the production of indolocarbazoles
XX      in Streptomyces (especially Streptomyces albus) using rebeccamycin
XX      CDS biosynthetic genes isolated from Saccharothrix aerocolonigenes ATCC39243.
XX      The invention also relates to the Saccharothrix aerocolonigenes ATCC39243
XX      rebeccamycin biosynthetic gene cluster (ADJ53165) and sequences at least
XX      80% homologous to it; nucleic acids encoding one or more rebeccamycin
XX      CDS biosynthetic proteins; the 18 rebeccamycin biosynthetic proteins
XX      CDS (ADJ53166-ADJ53183) encoded by the gene cluster; and vectors and host
XX      cells comprising nucleic acid sequences of the invention.
XX      Indolocarbazoles (especially rebeccamycin) and their derivatives and
XX      precursors are useful as antitumor agents. Rebeccamycin also shows
XX      antibacterial activity against Gram positive bacteria such as
XX      Staphylococcus aureus, Micrococcus luteus and Streptococcus faecalis. The
XX      recombinant Streptomyces albus of the invention produce increased yields
XX      of indolocarbazoles compared with Saccharothrix aerocolonigenes
XX      ATCC39243, and may permit the production of novel indolocarbazole
XX      compounds (e.g., halogenated derivatives). The present sequence
XX      represents the specifically claimed rebeccamycin biosynthetic gene
XX      cluster from Saccharothrix aerocolonigenes ATCC39243.
XX
XX      Sequence 25681 BP; 4040 A; 9104 C; 8809 G; 3728 T; 0 U; 0 Other;
SQ

```

Query Match 17.1%; Score 141.8; DB 10; Length 25681;  
 Best Local Similarity 55.9%; Pred. No. 6.7e-21;  
 Matches 291; Conservative 0; Mismatches 227; Indels 3; Gaps 1;

QY 16 GGACCAACATCCGAGAGGTTGGGCAAGATATGACTGTGACCGCTTGTGAACTCG 75  
 |||||  
 DB 16584 GCAACGACCCCTGAGAGAGTACGAGGAGATGTACGAGACTTCAACCGACCTTTCGCAAG 16643  
 |||||  
 QY 76 GTCCGCGGCGGCGCCCTGCGCCATCCACACGCGCTACTGTGGAGAACGACGGGCGGCTTCC 135  
 |||||  
 DB 16644 ATCTGGGGGGAGAACTGTGACTTTCGGCTACTGGGA---GGACGCGGCGCCGACGCTTCC 16700  
 |||||  
 QY 136 TGGCAGCAGAGCCCGCAGCCGCTTACCGACCTTTCGCGGAAAGACGCTGTCTGATGAGC 195  
 |||||  
 DB 16701 GTCCAGCAGCCAGCAGGAGCGGCTGTACCGACGAGATATGAGCCCTGCTCGACGCTCCGATCA 16760  
 |||||  
 QY 196 GGGCTTCCAGCTCTGATGATGGGGTGGCGGTACCGGACCAACGCGCTGGCGCTCGCGCC 255  
 |||||  
 DB 16761 GGGGATCGGGTCTGAGCGTCTGCGGATCGGAGATCGGACCCGCGCTGGCGCTCGCACG 16820  
 |||||  
 QY 256 GACAAAGCGGATCCAGATCAACCGGACATCCGTCAGCAGGTCGAAAGTGCGCATCGCCGCT 315  
 |||||  
 DB 16821 GCGAGGAGAGTCAAGGATGACAGGCACTTCGATCAGAGGCGCGAGTGAACAGGCGCAAC 16880  
 |||||  
 QY 316 GATTGCGCAGCGGAAACCGGACTTAAGCCACCGGATGAGTCTTCTGCTGCTGATGCCATG 375  
 |||||  
 DB 16881 GCGCAGCAGCCAGCGCGCGGCTGTGGCCAAACCGGATGACGCTTCTGATCCGACCGATG 16940  
 |||||  
 QY 376 TCCCTGCGGTACCCGGAACAATGCTTTCAGCGCGCGCTGGGCGCATGAGTCCGTGTGAG 435  
 |||||  
 DB 16941 GACCTGCGCTTGAAGAGCCGCTTCCTTCAGCGGAGTGTGGGCGCTGAGTGCCTGACAC 17000  
 |||||  
 QY 436 ATGTCCGACCGGACCGTGCCTATCCGGAATCTTTCAGTACTCAAAACCGGTGAGCATC 495  
 |||||  
 DB 17001 ATGCCGAGACCGGCGCGCGCACTTTCGCGAGATGGACAGCGGCTCTCGCGCGGATGGAC 17060  
 |||||  
 QY 496 CTCGCGGTACCGAGGTCGTCAAAAGAGAGCGCGCGCGCG 536  
 |||||  
 DB 17061 GTCCGATCGCGAGACTTCTGTCTGCTCGCGCTGTTCGAGG 17101  
 |||||

RESULT 15  
 AAC55802  
 ID AAC55802 standard; DNA; 852 BP.  
 XX  
 AC AAC55802;  
 XX  
 DT 19-JAN-2001 (first entry)  
 XX  
 DE S. lavendulae Mitw encoding DNA sequence.  
 XX  
 KW Mitomycin; biosynthesis; mtosane ring system; antibiotic; anti-cancer;  
 KW anti-inflammation; immune-enhancer; immunosuppressant; asthma;  
 KW chronic obstructive pulmonary disease; respiratory inflammation;  
 KW fungicide; pesticide; ds.  
 XX  
 OS Streptomyces lavendulae.  
 XX  
 PN WO200053737-A2.  
 XX  
 PD 14-SEP-2000.  
 XX  
 PF 10-MAR-2000; 2000WO-US006394.  
 XX  
 PR 12-MAR-1999; 99US-00266965.  
 XX  
 PA (MINU ) UNIV MINNESOTA.  
 PA (SHER) SHERMAN D H.  
 PA (MAOY) MAO Y.  
 PA (VARO) VAROGLU M.  
 PA (HEM) HE M.  
 PA (SHEL) SHELTON P C.

PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;  
 DR WPI; 2000-601980/57.  
 XX P-PsDB; AAB32457.  
 PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene  
 PT cluster useful for cloning mitomycin biosynthetic genes for elucidating  
 the molecular basis of mtosane ring system biosynthesis.  
 PS Example 1; Page 262-263; 399p; English.  
 XX  
 CC This invention relates to isolated and purified nucleic acid molecules  
 CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of  
 CC natural products that contain a variety of functional groups, including  
 CC amino benzquinone and axiridine ring systems. The S. lavendulae  
 CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes spanning  
 CC 55kb of DNA. The invention includes an expression cassette comprising a  
 CC mitomycin biosynthetic gene operably linked to a promoter, and host cells  
 CC transformed with the cassette. The nucleotide, and protein sequences and  
 CC the transformed host cells of the invention result in antiasthmatic,  
 CC antiinflammatory, cytosratic, immunomodulatory, and antibiotic  
 CC activities. The nucleotide sequences are used to elucidate the molecular  
 CC basis for the biosynthesis of the mtosane ring system, as well as to  
 CC engineer the biosynthesis of novel natural products, e.g. antibiotics,  
 CC anti-inflammatory agents, anti-cancer agents, immune-enhancers,  
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary  
 CC disease as well as other disease involving respiratory inflammation, or  
 CC cholesterol-lowering agents or as crop protection agents (e.g. fungicides  
 CC or insecticides) as well as biopolymers, e.g., in packaging or biomedical  
 CC applications, or to engineer PHA monomer syntheses. Sequences AAC5782-  
 CC C55811, AAC55815-C55849 and AAB32485-B32542 represent mitomycin  
 CC biosynthetic gene cluster DNA sequences and encoded proteins. Sequences  
 CC AAC55812-C55814, AAC55850-C55856 and AAC55862-C55869 represent PCR  
 CC primers used in the cloning of the mitomycin biosynthetic genes  
 XX  
 SQ Sequence 852 BP; 125 A; 316 C; 286 G; 125 T; 0 U; 0 Other;

Query Match 16.1%; Score 133.2; DB 3; Length 852;  
 Best Local Similarity 54.0%; Pred. No. 3.8e-19;  
 Matches 295; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

QY 15 GCGACCAACATCCGAGAGGTTGGGCAAGATATGACTGTGACCGCTTGTGAACTC 74  
 |||||  
 DB 36 CTCACCGCGCTTCGAGAGAGTGGCGCGCTTACGACCGGTTCACCGGCTGGAGCGCGC 95  
 |||||  
 QY 75 GGTCCGCGGCGGCGCCCTGCGCCATCCACAGCGGCTACTGGAGAAAGACGAGGCGGCTTC 134  
 |||||  
 DB 96 CTCCTCGGCGAGAACTTCGACTTTCGCTACTGGAGT---CCCCGACAGCCAGGTGCC 152  
 |||||  
 QY 135 CTGCGAGAGGCGCGCCGACCGGCTTACCGACCTTTCGCGGACGAGACCGGTTCATGG 194  
 |||||  
 DB 153 GCTGGCGAGGCGACCGACCGGCTTACCGACATGATGAGCGGAGCGGCTGCGCATCGGCGC 212  
 |||||  
 QY 195 CGGCGTTCGATGCTGCTGATGATGGGCTGCGGATCCGAGACCAACGACGCTGCGCGCGG 254  
 |||||  
 DB 213 CGGCTCCGCGCTCTGAGACTTCGCGGCTGGGAGTGGGAGACCCCGGCGCTGACGATCGCGC 272  
 |||||  
 QY 255 CGACAAAGCGGATCCAGATCAACCGGATCAACCGGATCAACCGGATCAACCGGATCAACCGG 314  
 |||||  
 DB 273 GCTCAGCGAGAGGAGATGCTCAGGAGATCTCGGTAGGACATAGAGGTCTGTCGGGCGCA 332  
 |||||  
 QY 315 TGATTGCGCAGCGAGAGCGGACTTAAGCAACCGGATGAGCTTTCGTCGCTGATGCGCAT 374  
 |||||  
 DB 333 CGGCTGGCGAGAGAGCGGCGCTTCGCGACCGGCGCGCTTCAGCGGCGCGACGCGAT 392  
 |||||  
 QY 375 GTCCCTGCGGATCCCGGACCAATGCTTTCAGAGCGCGCTTGGGCGCATGATGCTGTTGA 434  
 |||||  
 DB 393 GGAACCTCCCTTTCAGAGAGCGAGACTTCGACGCGGATGATGCGCTTCGATATGATCATCA 452  
 |||||  
 QY 435 GATGTCGGAACCGGAGCGGTCATCCGGAATCTTCGATGATCAAAACCGGATGCGAT 494  
 |||||  
 DB 453 CATGCCGAGACCGCGCCAGAGTGTCTGCCCAAGTGGCGCGGCTGCGCGCGCGGCGCG 512  
 |||||

Qy	495	CCTCGGCGCTCACCAGAGTGTCTCAACGAGAGCGGGCGGGGATGCCGTGTCGGGGA	554
Db	513	TCTGTGCTCACCGACTTCTCGAGCGGGCCCCCTCTGCCCCGAGGGCGGGCGCCGT	572
Qy	555	CAGTG	560
Db	573	CCAGCG	578

Search completed: August 29, 2005, 15:06:02  
 Job time : 593 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 29, 2005, 14:40:45 ; Search time 3416 Seconds  
(without alignments)  
9226.351 Million cell updates/sec

Title: US-10-069-353a-7

Sequence: 1 gctgtccagatgctgcgcacc.....cgcggaagcgcgtcgcgcga 828

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hcc:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gsa1:\*  
9: gb\_gsa2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88.4	10.7	1016	7 CK207033	CK207033 FGAS01864
2	87.8	10.6	695	5 BQ838139	BQ838139 WHE2907_A
3	84.4	10.2	787	6 CA764573	CA764573 AP53-RP1
4	82.6	10.0	587	7 CN008962	CN008962 WHE2647_B
5	80.6	9.7	878	7 CK196136	CK196136 FGAS00458
6	80.4	9.7	492	1 AV624709	AV624709 AV624709
7	80.4	9.7	680	4 BG858904	BG858904 1024060C0
8	80.4	9.7	730	4 BG861169	BG861169 1024076A0
9	80.4	9.7	803	4 BG847555	BG847555 1024017H1
10	79.8	9.6	571	2 BF257939	BF257939 HVSMEF001
11	79.8	9.6	618	6 CB866921	CB866921 HCO9M21W
12	79.8	9.6	622	6 CA210798	CA210798 SCEPSR113
13	79.8	9.6	666	6 CD442528	CD442528 EL01N0412
14	77.6	9.4	874	9 CC717854	CC717854 OGTA107C
15	77.6	9.4	908	9 CC717850	CC717850 OGTA107C
16	76.8	9.3	912	5 CC720063	CC720063 OGUER12TV
17	76.8	9.3	427	5 BQ823320	BQ823320 1030108C1
18	74.4	9.0	554	9 CC623301	CC623301 OGUHV56TH
19	74.4	9.0	638	9 CC685816	CC685816 OGMHUC8TV
20	74.2	9.0	1113	7 CK217823	CK217823 FGAS02982
21	74	8.9	813	7 CF884047	CF884047 L1C037XJ
22	71.6	8.6	610	4 B1531904	B1531904 1024117D1
23	71.6	8.6	653	4 B1531362	B1531362 1024113C0
24	70.6	8.5	738	7 C0131978	C0131978 GR_EB441

25	70.6	8.5	799	7 C0118507	C0118507 GR_EB020
26	70.6	8.5	855	7 C0076561	C0076561 GR_Ea37P
27	70.6	8.5	866	7 C0116523	C0116523 GR_EB018
28	69.8	8.4	705	7 C0435133	C0435133 7d5-c3 Mo
29	69	8.3	794	7 C0091150	C0091150 GR_Ea11H
30	68.8	8.3	588	2 BF619597	BF619597 HVSMEC000
31	68.8	8.3	719	5 BQ838223	BQ838223 WHE2907_H
32	68.8	8.3	781	6 CA405371	CA405371 9 A. brax
33	67.2	8.1	814	5 CN812339	CN812339 Fg13_0192
34	66.8	8.1	414	1 AV638404	AV638404 AV638404
35	65.8	7.9	714	6 CB970276	CB970276 CAB10003
36	65.6	7.9	742	7 C0524520	C0524520 3530_1_16
37	65.4	7.9	740	2 BE602586	BE602586 HVSMEH010
38	64.6	7.8	696	7 C0435474	C0435474 11f12_F12
39	64.6	7.8	699	7 C0435049	C0435049 6c7_CAS-0
40	64.4	7.8	311	6 C91811	C91811 C91811 Rice
41	64.4	7.8	541	5 BQ134030	BQ134030 samsd04
42	63.6	7.7	777	7 C0127869	C0127869 GR_EB12K
43	63.6	7.7	780	7 C0124614	C0124614 GR_EB07K
44	63.6	7.7	792	7 C0131974	C0131974 GR_EB441
45	63.6	7.7	802	7 C0123977	C0123977 GR_EB061

#### ALIGNMENTS

RESULT 1  
CK207033  
LOCUS  
DEFINITION  
aestivum CDNA, mRNA sequence.  
ACCESSION  
CK207033  
VERSION  
CK207033.1  
KEYWORDS  
SOURCE  
ORGANISM  
EST.  
Triticum aestivum (bread wheat)  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
1 (bases 1 to 1016)  
Allard, F., Crosby, W.L., Danyluk, J., Budes, F., Frick, M., Gaudet, D.,  
Genewein, B., Graf, R., Glick, P., Hrycan, L.D., Iaroché, A.,  
Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,  
Penkler, C., Roach, J.L. and Sarhan, F.  
Functional Genomics of Abiotic Stress In Wheat and Canola Crops  
Unpublished (2003)  
Contact: Wm L Crosby  
Bioinformatics  
University of Saskatchewan, Department of Computer Science  
1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033  
Email: fgas\_est@cs.usask.ca  
This sequence is the direct result of the Base calling software  
phred (default parameters). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region [31,821].  
Place: L5B009 row: G column: 22.  
Location/Qualifiers  
1..1016  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="Triticum aestivum FGAS: Library 5 GATB 7"  
/note="Vector: PCMW.SPORT6; Crown and developmental stages  
of spike formation in wheat cultivar Norstar. 4 mRNA  
populations were combined before constructing the library.  
The first mRNA population is from 1cm crown sections after  
30 days of cold acclimation. The second is from 1cm crown  
sections after 11 days of deacclimation (before  
deacclimation plants were fully vernalized for 49 days).

TITLE  
JOURNAL  
COMMENT

#### FEATURES

source

The third is from different developmental stages of spike formation (5 to 50mm) that still have not emerged from the leaf (dissection required). The last is from different developmental stages of spike and seed formation after having emerged from the leaf (visible). First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI."

# ORIGIN

Query Match 10.7%; Score 88.4; DB 7; Length 1016;  
Best Local Similarity 52.0%; Pred. No. 1.2e-10;  
Matches 197; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

139 CAGCAGCGCCGCGGCTCAACGACCTTGTGCGCAACGAGCGTGTGATGCGGC 198  
407 CGGAGGCGGCGGCGGCTGACGAGGCGCGGCGGCGGCGGCGGCGGCGGCGG 466  
199 GTTCACTGTCTGATGTGGGCTGCGGTAACGAGCAACGCGCTGCGGCGGCGG 258  
467 AAGCGGCTGCTGACGTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 526  
259 AAGCGGCTGCTGACGTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 318  
527 TCCGAGCTGCGAGTGTGCGGATCAACGAGTACGAGTACGAGTACGAGTACG 586  
319 TCCGAGCTGCGAGTGTGCGGATCAACGAGTACGAGTACGAGTACGAGTACG 378  
587 CACAAACCGGAGGCGGCGGCTGACGCGCAATGCGAGTGTGCGGCACTTCAATG 646  
379 CTGCGGTAACCGGAGCAATGCTTTCAGACGCGGCTGCGGCGGCGGCGGCGG 438  
647 ATGCGGTTCCCGAGCGGCGGCTTTCAGACGCGGCTTTCAGACGCGGCGGCGG 706  
439 TCCGAGCTGCGAGTGTGCGGATCAACGAGTACGAGTACGAGTACGAGTACG 498  
707 CCGAGGCTGCGAGGAGTGTGCGGATCAACGAGTACGAGTACGAGTACGAGT 766  
499 GCGGTACCGGAGGAGTGTGCGGATCAACGAGTACGAGTACGAGTACGAGT 517  
767 GTCTCTACGAGTGTGCGGATCAACGAGTACGAGTACGAGTACGAGTACGAG 785

RESULT 2 695 bp mRNA linear EST 08-AUG-2002  
B0838139  
LOCUS B0838139  
DEFINITION WHE2907\_A04\_B07S wheat aluminum-stressed root tip cDNA library  
Triticum aestivum cDNA clone WHE2907\_A04\_B07, mRNA sequence.

ACCESSION B0838139  
VERSION B0838139.1 GI:22142457  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum

REFERENCE 1 (bases 1 to 695)  
AUTHORS Anderson,O.D., Chao,S., Chin,A., Close,T.J., Gustafson,J.P.,  
Lazo,G.R., Rausch,C.J., Ross,K., Seaton,C.L. and Wilson,C.  
The structure and function of the expressed portion of the wheat  
genomes - Aluminum-stressed root tip cDNA library  
Unpublished (2001)

JOURNAL COMMENT  
US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: oanderson@pw.usda.gov  
Sequences have been trimmed to remove vector sequence and low  
quality sequence with phred score less than 20  
Seq primer: SK primer.  
Location/Qualifiers

FEATURES

## source

1..695  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="BHL146"  
/db\_xref="taxon:4565"  
/clone="WHE2907\_A04\_B07"  
/tissue\_type="Root tip at 1.0 to 1.5 mm stage"  
/dev\_stage="Seedling"  
/lab\_host="E. coli SOLR"  
/clone\_lib="Wheat aluminum-stressed root tip cDNA library"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid.  
Site\_1: EcoRI; Site\_2: XhoI; Plants were grown under  
hydroponic conditions, root tips were excised and snap  
frozen, total RNA was prepared at University of  
Missouri (Ross, Gustafson). Poly(A) RNA was purified, a  
cDNA library was made, and the cDNA clones were in vivo  
excised to give phagescript SR-phagemids in the T7 close  
lab (Chin and Close) at the University of California,  
Riverside. Plasmid DNA preparations and DNA sequencing  
were performed in the OD Anderson lab (all other  
authors)."

## ORIGIN

Query Match 10.6%; Score 87.8; DB 5; Length 695;  
Best Local Similarity 52.0%; Pred. No. 1.6e-10;  
Matches 197; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

139 CAGCAGCGCCGCGGCTCAACGACCTTGTGCGCAACGAGCGTGTGATGCGGC 198  
261 CGGAGGCGGCGGCGGCTGACGAGGCGCGGCGGCGGCGGCGGCGGCGGCGG 320  
199 GTTCACTGTCTGATGTGGGCTGCGGTAACGAGCAACGCGCTGCGGCGGCGG 258  
321 AAGGCGCTGCTGACGTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 380  
259 AAGCGATCCAGTACCGGCGGATCAACGAGTACGAGTACGAGTACGAGTACG 318  
381 TCCGAGCTGCGAGTGTGCGGATCAACGAGTACGAGTACGAGTACGAGTACG 440  
319 TCCGAGCTGCGAGTGTGCGGATCAACGAGTACGAGTACGAGTACGAGTACG 378  
441 CACAAACCGGAGGCGGCGGCTGACGCGCAATGCGAGTGTGCGGCACTTCAATG 500  
379 CTGCGGTAACCGGAGCAATGCTTTCAGACGCGGCTGCGGCGGCGGCGGCGG 438  
501 ATGCGGTTCCCGAGCGGCGGCTTTCAGACGCGGCTTTCAGACGCGGCGGCGG 560  
439 TCCGAGCTGCGAGTGTGCGGATCAACGAGTACGAGTACGAGTACGAGTACG 498  
561 CCGAGGCTGCGAGGAGTGTGCGGATCAACGAGTACGAGTACGAGTACGAGT 620  
499 GCGGTACCGGAGGAGTGTGCGGATCAACGAGTACGAGTACGAGTACGAGT 517  
621 GTCTCTACGAGTGTGCGGATCAACGAGTACGAGTACGAGTACGAGTACGAG 639

RESULT 3 787 bp mRNA linear EST 08-JAN-2003  
CA764573  
LOCUS CA764573  
DEFINITION AF33-RP6\_04\_B21.T7\_081.ab1 IRR1 Drought Stress Panicle Library  
Oryza sativa (indica cultivar-group) cDNA clone C0001197 5' similar  
to Delta(24)-sterol C-methyltransferase (BC 2.1.1.41), mRNA  
sequence.

ACCESSION CA764573  
VERSION CA764573.2 GI:27546528  
KEYWORDS EST.  
SOURCE Oryza sativa (indica cultivar-group)  
ORGANISM Oryza sativa (indica cultivar-group)  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehharoidae; Oryzaeae; Oryza.

REFERENCE 1 (bases 1 to 787)  
AUTHORS Bennett,J., Arumugam,K., Lafitte,R., Wen,J., Rudd,S. and

TITLE Bruskiewicz, R.M.  
JOURNAL IRRI Drought Stress Panicle cDNA Library  
COMMENT Unpublished (2002)  
On Dec 2, 2002 this sequence version replaced gi:25993828.

CONTACT: Richard Bruskiewicz  
Biometrics and Bioinformatics Unit  
International Rice Research Institute  
DAPO 7777, Metro Manila, Philippines  
Tel: +63-2-845-0563  
Fax: +63-2-845-0606  
Email: r.bruskiewicz@cgiar.org  
International Rice Information System (IRIS);  
http://www.iris.irri.org; D0201196  
Assignment of putative function to the sequence by S. Rudd of the  
Munich Information Center for Protein Sequences  
(http://mips.gsf.de)  
Place: 04 row: B column: 21.  
Location/Qualifiers

## FEATURES

1..787  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="IR64"  
/db\_xref="taxon:39946"  
/clone="C0001197"  
/issue\_type="Panicles"  
/dev\_stage="flowering"  
/clone\_lib="IRRI Drought Stress Panicle Library"  
/note="Vector: pBluescript II SK+; Water stresses was applied by not watering for 4 consecutive days. Panicles were collected from control (well watered) and stressed plants at 2 days before heading, at heading, 50% flowering and 4 days after 50% flowering."

## ORIGIN

Query Match 10.2%; Score 84.4; DB 6; Length 787;  
Best Local Similarity 51.0%; Pred. No. 1.1e-09;  
Matches 199; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

QY 128 GGGCTTCTGGAGCAGCGCCGCGACCGGCTCAACCACTTGTGCGCAACGAGACCGTGC 187  
DB 121 GAGAGTCTCTTATCAAGGCGCATTCGCGCATAGACATACTGCGCTGCATGATTGGCA 180  
QY 188 TCGATGCGCGCGCTTCACTGCTGATGTGGGGTGGCGGTACCGAACAACGCGCTCGGG 247  
DB 181 TCAAGAGGGCATGAAGGTTCTGATGTGGGCTGGGTTGGTGTCTGACGCTCGAGAGA 240  
QY 248 TCGCGCGCGCAACGCGCATCAAGATCAACCGCATCAACCGTCAAGTGCAGAGTGC 307  
DB 241 TTGCCAAGTTCAACCGGTGCCCATTAACCGGCTCAACAACAAGCATATCAATCAAG 300  
QY 308 TCGCGCGTGAATTCGCAACGCGGAACCGGACTAAGCAACCGGTTGACTTCTCGTGGTG 367  
DB 301 GTGCTACCCACTACGCGCAAGAACAGAGAGCTCTCTCAACGATTTGTCAAGGGCG 360  
QY 368 ATGCCATGTCCTCGCGTACCGGCAATGCTTTCAACCGCGCGCTGGCGCATGACGCG 427  
DB 361 ACTTCATGACAGATGCTCTTCCTGACAACTCTTCGACGCTGTGACGCCATGAGGCTA 420  
QY 428 TGTGAGAGATGTCGAACCGGACCGTGCATTCGCGGGAATCTTCGAGTACTGAACCCG 487  
DB 421 CCTGCCAGCGCCCAACCGTTCGAGAGGCAATTAACGCGAGATCTTCGATGCTCAAGCCG 480  
QY 488 GTGGCAATCTCGGCGTCAACCGAGTCTCA 517  
DB 481 GCGGCGTCTTCGGTGTATTACAGAGGCTCA 510

RESULT 4  
CN008962 587 bp mRNA linear EST 29-MAR-2004  
LOCUS CN008962  
DEFINITION WHE2647\_B12\_D23Z Wheat Fusarium graminearum infected spike cDNA  
library Triticum aestivum cDNA clone WHE2647\_B12\_D23, mRNA  
sequence.

ACCESSION CN008962  
VERSION CN008962.1 GI:45798994  
KEYWORDS EST  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum

REFERENCE  
AUTHORS Anderson, O.D., Chao, S., Han, P.S., Heinen, S., Hata, C.C., Kang, Y., Krueger, W.M., Lazo, G.R., Miller, S., Muehlbauer, G.J., Miller, R., Pritsch, C., Rausch, C.J., Seaton, C.L., Tong, J.C., Vance, C. and Wilson, C.F.  
1 (bases 1 to 587)  
The structure and function of the expressed portion of the wheat  
genomes - Fusarium graminearum infected spike cDNA library  
Unpublished (2001)

JOURNAL  
COMMENT The structure and function of the expressed portion of the wheat  
genomes - Fusarium graminearum infected spike cDNA library  
Unpublished (2001)  
Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: oanderson@pw.usda.gov

Sequences have been trimmed to remove vector sequence and low  
quality sequence with phred score less than 20. No effort was taken  
to identify ESTs of fungal origin from this library, thus this EST  
could be of wheat or fungal origin.  
Seq primer: T3 primer.

## FEATURES

## source

1..587  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Suma13"  
/db\_xref="taxon:4565"  
/clone="WHE2647\_B12\_D23"  
/issue\_type="Spike"  
/dev\_stage="Adult plant"  
/lab\_host="E. coli S0LR"  
/clone\_lib="Wheat Fusarium graminearum infected spike cDNA  
library"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid  
pBluescript SK, Site 1: EcoRI, Site 2: XhoI; Plants were  
grown in the greenhouse. Spikes were sprayed at anthesis  
with Fusarium graminearum. Total RNA, and poly(A) RNA were  
prepared and pooled from infected spike at 0, 6, 12, 24,  
36 and 48 hours after inoculation, a cDNA library was  
made, and the cDNA clones were in vivo excised to give  
pBluescript phagemids in G. Muehlbauer lab at the  
University of Minnesota (Krueger, W.M., Muehlbauer, G.J.,  
Pritsch, C., Vance, C.). The cDNA library should contain  
genes of both wheat and fungal pathogen origin. Plasmid  
DNA preparations and DNA sequencing were performed in the  
OD Anderson lab (all other authors)."

## ORIGIN

Query Match 10.0%; Score 82.6; DB 7; Length 587;  
Best Local Similarity 50.4%; Pred. No. 2.9e-09;  
Matches 202; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

QY 134 CCTGGCAGACCGCGCGACCGGCTCAACCACTTGTCCGCAACGAGACCGTGTGATG 193  
DB 15 CTTTATCAAGGCGCATGTCGGTCAATGACACTTGTCCCAACGAGTGGATCAAG 74  
QY 194 GCGGCGTGTGACTGCTCGATGTGGGTGGGTGCAACGAGACCGAGCGTGGCGCC 253  
DB 75 ACGGCATGAAGGTTCTTGATGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 134  
QY 254 GCGACAAAGCGATTCAGATCAACCGCATCAACCGTCAACGAGTGAAGTGGCATGCGCG 313  
DB 135 AGTTCAAGAGTGTCTATTTACCGGCTTAACAACAACAATCAACAGATCGAGGCTA 194  
QY 314 CTGATTGCGCAACGAGACCGGATCAAGCCACCGGATGACTTCTCGTGGTGTGATGCA 373

Db 195 CCCACTGACCTTCAAGAGGGTCTGTCCAACCACTTGATTGTTCAGAGGTGATTCA 254  
 QY 374 TGTCCCTGCGCTACCGGCAATGCTTTGAGCGCGGCTGAGGCGCATGAGTGCCTGG 433  
 Db 255 TGCAAATGCTTTCCCGCAACAGCTTGAGTGGTTATGCAATGAGGCGACCGTTC 314  
 QY 434 AGATGTCGCAACCGGACCGTGCATCCGGAAATCTTCAGTACTCAACCCGGTGC 493  
 Db 315 AGGCTCCGACCTCAAGGGCATCTACAGGAGATTTTCGTCTTTAAGCCCGGTGG 374  
 QY 494 TCTCTGGCGCTCACCGAGTGTCAACGAGAACGGCGCGC 534  
 Db 375 TCTCGGTGTGTACGAGTGGCTCATGACTACGAGTACGAC 415  
 RESULT 5  
 LOCUS CK196136 878 bp mRNA linear EST 08-DEC-2003  
 DEFINITION FGAS004583 Triticum aestivum FGAS: Library 3 Gate 6 Triticum  
 accession cDNA, mRNA sequence.  
 CK196136  
 VERSION CK196136.1 GI:39558526  
 KEYWORDS EST  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooideae; Triticeae; Triticum.  
 1 (bases 1 to 878)  
 Allard, F., Crosby, W. L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
 Genswein, B., Graf, R., Gulick, P., Hrycan, L. D., Laroché, A.,  
 Links, M. G., McCarthy, E. L., Monroy, A., Muzak, I., Nilsson, D.,  
 Peniket, C., Roach, J. L. and Sarhan, F.  
 Functional Genomics of Abiotic Stresses In Wheat and Canola Crops  
 Unpublished (2003)  
 CONTACT: Wm L Crosby  
 Bioinformatics  
 University of Saskatchewan, Department of Computer Science  
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
 Saskatchewan, S7N 5A9, Canada  
 Tel: 306 966 1769  
 Fax: 306 966 2033  
 Email: fgas\_est@cs.usask.ca  
 This sequence is the direct result of the Base calling software  
 Phred (default parameters). It is the raw base calls. To aid in the  
 identification of the high quality insert the software Lucy  
 (default parameters) has been run on this sequence. Lucy identified  
 the region [102,750].  
 Plate: L3C105 row: O column: 10.  
 Location/Qualifiers  
 1..878  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4565"  
 /clone\_lib="Triticum aestivum FGAS: Library 3 Gate 6"  
 /note="Organ: Root; Vector: PCWV.SPORT6; Root tissue from  
 control, cold-acclimated and salt stressed wheat cultivar  
 Norstar. 7 mRNA populations were combined before  
 constructing the library; 7 day non-acclimated roots, 1,  
 23, and 53 days cold-acclimated at 4C, and 30 minutes, 3  
 hours and 6 hours treated roots with 200mM NaCl.  
 Non-acclimated and cold-acclimated plants were grown in  
 vermiculite while salt stressed plant were grown  
 hydroponically. First strand synthesis in this library was  
 done in the presence of methylated dCMP thereby protecting  
 from internal cleavage with NotI."

ORIGIN  
 Query Match 9.7%; Score 80.6; DB 7; Length 878;  
 Best Local Similarity 51.4%; Pred. No. 8.9e-09;  
 Matches 185; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 133 TCTGCGACGACGCGCGGCTCAACCGACTTGTGCGCAAGGACCGTCTGAT 192

Db 499 TCCACGCGGACGACACCGCGCTCCAGAGAGCGCGTCCGCACTGTGGGGGCAAG 558  
 QY 193 GGGCGGCTTGCATCTGCTGATGTGGGGTGGCGGTACCGGACAAACAGCGGTCCGGCGG 252  
 Db 559 CCGGCGCACCGCGTGTGAGAGTGGCTGGCGGTGGCGGGCCCATATGCGGCATCGG 618  
 QY 253 CGGCACAACGCGATTCAGATCAACCGGATCAACGTCAGCCAGGTGAGTGGCCATCGCC 312  
 Db 619 GCGCATCCGGCGGCGCGGTGTGTGGCATACATCAACAGATACAGGTAAACGGGGCC 678  
 QY 313 GGTATGTGCGACGCGGACCGGACTAAGCCACCGGGTGAATTCTGTGCTGTGATGCC 372  
 Db 679 CGCGCGCAACACCGCAAGGCGGGCTGGAACCGCANTGCCAGGTGTGTGGCGCACTTC 738  
 QY 373 ATGTCCCTGCGCTACCGGCAATGCTTTGAGCGCGGCTGGGCGCATGAGTGGCTGTG 432  
 Db 739 ATGGCGATGCCCTTGTGACGAGCGCTCTTCAAGCGGCGCTACTCATGAGGCACTGC 798  
 QY 433 GAGATGTCCGAACCGGACCGGTGCATCCGGAAATCTTCAAGTACTCAACCCGCTGCG 492  
 Db 799 CACGCGCCGACGCTGACAGAGCTGTACGGGAGGTGTACCGGGTCTCATGCGGGCGC 858  
 RESULT 6  
 LOCUS AV624709 492 bp mRNA linear EST 15-DEC-2000  
 DEFINITION AV624709 Chlamydomonas reinhardtii 5' to 0.04% CO2 Chlamydomonas  
 reinhardtii cDNA clone LC081e06\_r 5', mRNA sequence.  
 AV624709  
 VERSION AV624709.1 GI:10773886  
 KEYWORDS EST  
 SOURCE Chlamydomonas reinhardtii  
 ORGANISM Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 Chlamydomonadales; Chlamydomonadales.  
 1 (bases 1 to 492)  
 Asamizu, E., Miura, K., Kuchino, K., Inoue, Y., Fukuzawa, H., Ohnaka, K.,  
 Nakamura, Y. and Tabata, S.  
 Generation of expressed sequence tags from low-CO2 and high-CO2  
 adapted cells of Chlamydomonas reinhardtii  
 DNA Res. 7 (5), 305-307 (2000)  
 JOURNAL MEDLINE 20539644  
 PUBMED 11089912  
 CONTACT: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
 Location/Qualifiers  
 1..492  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="C9"  
 /db\_xref="taxon:1055"  
 /clone\_lib="LC081e06\_r"  
 /note="Vector: pBluescriptII SK-; Site 1: BclRI; Site 2:  
 XhoI; The cDNA library was constructed from cells cultured  
 in a carbon stress acclimatized condition in which carbon  
 dioxide concentration in the bubbling gas was changed from  
 5% to 0.04%"

ORIGIN  
 Query Match 9.7%; Score 80.4; DB 1; Length 492;  
 Best Local Similarity 51.1%; Pred. No. 9.6e-09;  
 Matches 189; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 146 CCGCGCACCGGCTACCGGACTTGTGCGCGGACCGGCTGATGCGCGGCTTGAC 205  
 Db 53 CCGAGCGCGCGACGAGGCGGCGCATCGCTGCGTGTGCGGCGCGGCGCAAGG 112  
 QY 206 TGTCTGATGTGGGGTGGTACCGGACCAACGCGCTGCGCGCGGCAACCGGA 265





PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with Exsist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

## ORIGIN

Query Match 9.7%; Score 80.4; DB 4; Length 730;  
Best Local Similarity 51.1%; Pred. No. 9.9e-09;  
Matches 189; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

146 CCGCGACCGGCTCACCGGACCTTGTGCGGACGACCGTGTGCGGCGGCGTTCGAC 205  
|||  
10 CCGAGGCGCGGACGAGGCGCGCATCGTCCCTGCTGTGCGCTGCAAGCCGCGAGAG 69  
|||  
206 TGCTCGATGTGGGGTGGGTACCGGACCAACGACGCTGCGCGCGGACGACGCA 265  
|||  
70 CGCTGAGACTGCGGCTGGGTGGGTGGCGCGATGCGTCCGTGCTGTCAAGCGGCG 129  
|||  
266 TCACATATCACCGGATCACCGTCAAGTGGCCATCGCGCTGATTTGGCGAC 325  
|||  
130 CGCATATCACCGGATCACCATCAACGATGAGCGCGGCAAGACGACGACG 189  
|||  
326 GGAACGCGGATTAAGCCACCGGCTGACCTTCTGCGGCGATGTCCTGCGGT 385  
|||  
190 CACGCGAGGTGTGCTCGGTGACGACGCTTGTCCGCGGCGATTCATCAATGCGGT 249  
|||  
366 ACCCGACATGCTTTGACGCGCGCTGGGCGATGAGTGTGTTGAGATGTCGAC 445  
|||  
250 TCAAGGAGAACACCTTTCGACGCGGCTTATGAGGCGACCTGCCACGCGCCCAAGC 309  
|||  
446 CGGACCGTGGCCATCCGGGAAATCCTTCAAGTACTAAACCGGTGGCATCTCCGGGCTCA 505  
|||  
310 TGGAGCAGGTGTACGCGGAGATCTACCGGTCTCAAGCCGCGGACGATCACTTGTGTGT 369  
|||  
506 CCGAGGTGCT 515  
|||  
370 ACGAGTGGGT 379

RESULT 9 803 bp mRNA linear EST 29-MAY-2001  
BG847555  
LOCUS 1024017H12.y1 C. reinhardtii CC-1690, normalized, lambda Zap II  
DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.  
ACCESSION BG847555  
VERSION BG847555.1 GI:14228739  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonadales; Chlamydomonadales;  
1 (bases 1 to 803)  
Grossman, A., Davies, J., Federpiehl, N., Harris, E., Lefebvre, P.,  
McDermott, J.P., Silflow, C., Stern, D. and Suzuki, R.  
Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants; project phase 2  
Unpublished (2000)  
Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.  
Location/Qualifiers  
1. 803  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21g+

## FEATURES

source

/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, normalized, lambda Zap II"  
/note="Vector: pBluescript II SK-, Site 1: EcoRI, Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in YAP (acetate-containing) medium in the light, YAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with Exsist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

## ORIGIN

Query Match 9.7%; Score 80.4; DB 4; Length 803;  
Best Local Similarity 51.1%; Pred. No. 9.9e-09;  
Matches 189; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

146 CCGCGACCGGCTCACCGGACCTTGTGCGGACGACCGTGTGCGGCGGCGTTCGAC 205  
|||  
10 CCGAGGCGCGGACGAGGCGCGCATCGTCCCTGCTGTGCGCTGCAAGCCGCGAGAG 69  
|||  
206 TGCTCGATGTGGGGTGGGTACCGGACCAACGACGCTGCGCGCGGACGACGCA 265  
|||  
70 CGCTGAGACTGCGGCTGGGTGGGTGGCGGATGCGTCCGTGCTGTCAAGCGGCG 129  
|||  
266 TCACATATCACCGGATCACCGTCAAGTGGCCATCGCGCTGATTTGGCGAC 325  
|||  
130 CGCATATCACCGGATCACCATCAACGATGAGCGCGGCAAGACGACGACGACG 189  
|||  
326 GGAACGCGGATTAAGCCACCGGCTGACCTTCTGCGGCGATGTCCTGCGGT 385  
|||  
190 CACGCGAGGTGTGCTCGGTGACGACGCTTGTCCGCGGCGATTCATCAATGCGGT 249  
|||  
366 ACCCGACATGCTTTGACGCGCGCTGGGCGATGAGTGTGTTGAGATGTCGAC 445  
|||  
250 TCAAGGAGAACACCTTTCGACGCGGCTTATGAGGCGACCTGCCACGCGCCCAAGC 309  
|||  
446 CGGACCGTGGCCATCCGGGAAATCCTTCAAGTACTAAACCGGTGGCATCTCCGGGCTCA 505  
|||  
310 TGGAGCAGGTGTACGCGGAGATCTACCGGTCTCAAGCCGCGGACGATCACTTGTGTGT 369  
|||  
506 CCGAGGTGCT 515  
|||  
370 ACGAGTGGGT 379

## RESULT 10

BF257939 571 bp mRNA linear EST 22-OCT-2001  
HVSMEF0014E17f Hordeum vulgare seedling root EST library HVCNDA0007  
DEFINITION (Ectisolated and unstressed) Hordeum vulgare subsp. vulgare cDNA  
clone HVSMEF0014E17f, mRNA sequence.  
ACCESSION BF257939  
VERSION BF257939.2 GI:13119150  
KEYWORDS EST.  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooidae; Triticaceae; Hordeum.  
1 (bases 1 to 571)  
Wing, R., Close, T.J., Kleinholz, A., Wise, R., Begum, D., Frisch, D.,  
Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W.,  
Fenton, R.D., Oates, R. and Main, D.  
Development of a genetically and physically anchored EST resource  
for barley genomics: Morex unstressed seedling root cDNA library  
Unpublished (2001)  
On Nov 16, 2000 this sequence version replaced gi:11187052.

## JOURNAL

COMMENT

## TITLE

REFERENCE

AUTHORS

## TITLE

Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants; project phase 2  
Unpublished (2000)  
Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.  
Location/Qualifiers  
1. 803  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21g+

## JOURNAL

COMMENT

## FEATURES

source

Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: twing@clemson.edu  
Total hg bases = 319  
Seq primer: AATTAACTTCACTTAAGG  
High quality sequence stop: 480.  
Location/Qualifiers

## FEATURES

1..571  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Morex"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="HVSMEF0014R17f"  
/issue\_type="Seedling root"  
/lab\_host="TU121"  
/clone\_lib="Hordeum vulgare seedling root EST library  
HYCDNA0007 (Etiolated and untreasured)"  
/note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;  
Seeds were surface sterilized then germinated under axenic  
conditions in the dark at room temperature on filter paper  
with water, nystatin and ceftioxiime in covered  
crystallization dishes. Five-day old seedling roots were  
then harvested, total RNA was prepared, poly(A) RNA was  
purified, one primary unamplified cDNA library was made,  
and 1 million pfu were in vivo excised to give pluescript  
SK(-) cDNA phagemid. These steps were performed in the TU  
Close Laboratory at the University of California,  
Riverside (Choi, Close, Fenton). Phagemids were plated and  
picked at the Clemson University Genomics Institute (CUGI)  
(Begum, Palmer, Friesch, Atkins and Wing). Plasmid DNA  
preparations, DNA sequencing and sequence analysis were  
performed at CUGI (Wing, Yu, Friesch, Henry, Simmons,  
Oates, Rambo, Main). The sequence has been trimmed to  
remove vector sequence and contains a minimum of 100 bases  
of phred value 20 or above. For more details on library  
preparation and sequence analysis see  
http://www.genome.clemson.edu/projects/barley. To order  
this clone see http://www.genome.clemson.edu/orders Also  
see Close TJ, Wing R, Kleinbols A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomics. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/gspages/bgn/31/cover.html)"

## ORIGIN

Query Match 9.6%; Score 79.8; DB 2; Length 571;  
Best Local Similarity 51.2%; Pred. No. 1.3e-08;  
Matches 186; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

139 CACGAGGCGCGGACCGGCTTGTGCGGAGGACCGGCTGAGTGGGCGC 198  
149 CGGAGGCGGACCGGCTGTCAGAGGACCGGCTGCTCGAGGCGCGCGG 208  
199 GTTCGACTGCTCATGTGGGGTGGGATACCGGACCAACGCGCTGCGGCGCGAC 258  
209 AACCGGCTGTGATCGCTGGGCTGGGGGCGGCGGCGGCGGCGGCGGCGAC 268  
259 AACGCGATTCAGATCAACCGGATCAACCGGATCAACCGGATCAACCGGCGGCTGC 318  
269 TCCGGCTCCGACGCTCGGGATCACTATCAACGATCAACGATCAACCGGCGGCTGC 328  
319 TGGGACGCGGACCGGATCAACCGGATCAACCGGATCAACCGGATCAACCGGATTC 378  
329 CACGACCGGATCGGCGGCTGAGACCAATGCGAGGCTGTCGCGGCGGCGGCGGCGG 388  
379 CTGCGGATCCCGGACCAATGCTTTCGACGCGGCTGCGGCGGCGGCGGCGGCGGCGG 438  
Db 389 ATGCGCGTCCGACGCGGCTTTCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 448

QY 439 TCCGACCGGACCGGCTGATCCGGAATCTTTCAGTACTCAACCGGCTGATCTTC 498  
Db 449 CCGAGCTTTCGACGCTTACGCGGAGGCTTTCGAGCTCTCAACCGGCGGCGGCTTAC 508  
QY 499 GGC 501  
Db 509 GTC 511

## RESULT 11

CB868921 618 bp mRNA linear EST 03-JUL-2003  
LOCUS CB868921  
DEFINITION HC09M21w CH Hordeum vulgare cDNA clone HC09M21 5-PRIME, mRNA  
sequence.  
ACCESSION CB868921  
VERSION CB868921.1 GI:30070898  
KEYWORDS EST.  
SOURCE Hordeum vulgare  
ORGANISM Hordeum vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooidae; Triticeae; Hordeum.  
1 (bases 1 to 618)  
Varshney,R.K., Zhang,H., Burton,R., Stein,N., Langridge,P. and  
Granel,A.

## REFERENCE

AUTHORS Barley ESTs from coleoptile tissue  
JOURNAL Unpublished (2003)  
COMMENT Contact: Stein Nils  
Molecular Markers Group, Department Genbank  
Institute of Plant Genetics and Crop Plant Research (IPK)  
Corrensstr. 3, 06466, Gatersleben, Germany  
Tel: 039482-5522  
Fax: 039482-5595  
Email: stein@ipk-gatersleben.de  
Insert length: 618 Std Error: 0.00  
Plate: 9 row: M column: 21  
Seq primer: T7.

## FEATURES

1..618  
/organism="Hordeum vulgare"  
/mol\_type="mRNA"  
/cultivar="Sloop"  
/db\_xref="GABI:547323"  
/clone="HC09M21"  
/issue\_type="coleoptile"  
/dev\_stage="coleoptile, 3 days old"  
/lab\_host="DH10B"  
/clone\_lib="CH"  
/notes="Vector: pSPORT; Site 1: SalI (5'-end of cDNA);  
Site 2: NotI (3'-end of cDNA); Due to the cloning system  
used blue/white selection for recombinants is not 100%  
reliable. Average insert size is 1.3 kb."

## ORIGIN

Query Match 9.6%; Score 79.8; DB 6; Length 618;  
Best Local Similarity 52.9%; Pred. No. 1.4e-08;  
Matches 171; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

195 CGGCGTTCGACTGCTGATGTGGGGTGGGATACCGGACCAACGCGCTGCGGCGG 254  
Db 3 CGGGAAGCGGCTCTGAGCTGCGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62  
QY 255 CGACGACCGGATTCAGATCAACCGGATCAACCGGATTCAGATTCAGATTCAGATTC 314  
Db 63 GCACTCGGCTCCGACGCTGCGGATCACTATCAACGATCAACGATCAACCGGCGGCGG 122  
QY 315 TGATTGCGACCGGACCGGATCAACCGGATCAACCGGATTCGCTGCGGATTCGATTC 374  
Db 123 CTGCAACGCGGACCGGCGGCTGAGCAAGCAATCGAGGCTGTCGCGGCGGCGGCGG 182  
QY 375 GTCCGCGGATCCCGGACCAATGCTTTCGACGCGGCTGCGGCGGCGGCGGCGGCGG 434

Db 183 GGCATGCGCTTCCGAGACGCGCTCTTCGACGCGCCTACTCCATCGAGGCCACCTGCCA 242

QY 435 GATGTCGGAACCGGACCGGTCATCCGGAATCCTTGAGTACTCAAAACCGGATGAT 494

Db 243 CGCGCCCAAGCTTCAGACGACGTTCACGCGGAGTCTTCGCGTCTCAAGCCAGGGGGCT 302

QY 495 CCTGGCGCTCACCGAGGTGTCGA 517

Db 303 CTACGCTTCATACGAGTGGGTCA 325

RESULT 12

CA210798 622 bp mRNA linear EST 25-SEP-2003

LOCUS SCEPSB1132A04.g SBI Saccharum officinarum cDNA clone SCEPSB1132A04

DEFINITION 5', mRNA sequence.

ACCESSION CA210798

VERSION CA210798.1 GI:35254474

KEYWORDS EST.

SOURCE Saccharum officinarum

ORGANISM Saccharum officinarum

REFERENCE 1 (bases 1 to 622)

AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.

TITLE The libraries that made SUCEST

JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

COMMENT Centro de Biologia Molecular e Engenharia Genetica Universidade Estadual de Campinas Caixa Postal 6010, 13083-970, Campinas SP, Brazil

Tel: 55 19 3788 1137

Fax: 55 19 3788 1089

Email: pattuda@unicamp.br

Clone distribution: clone distribution information can be found through the Brazilian Clone Collection Center (BCCC) at <http://www.bcccenter.fcav.unesp.br>

Plate: 132 row: A column: 04

Seq primer: T7 Promoter Primer.

Location/Qualifiers

1. 622

/organism="Saccharum officinarum"

/mol\_type="mRNA"

/db\_xref="taxon:4547"

/clone="SCEPSB1132A04"

/lab\_host="DH10B"

/clone\_1lb="SBI"

/note="Organ: Stalk Bark from adult plants; Vector: pSport1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from [Stalk Bark from adult plants]. cDNA was prepared from polyA+ mRNA using Superscript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a separose C4-28 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at <http://sucest.fad.ic.unicamp.br/public>"

ORIGIN

Query Match 9.6%; Score 79.8; DB 6; Length 622;

Best Local Similarity 52.1%; Pred. No. 1.4e-08;

Matches 202; Conservative 0; Mismatches 183; Indels 3; Gaps 1;

QY 133 TCTTGCGACAGCGCGCGCTCACCGACTTGTGCGGAAACGACCGTGTGAT 192

Db 137 TCCACACCGGACGCGCACGCGCTCCACGAGAGCGGTGCGGACCTCTCGGCCACAG 196

QY 193 GCGGCGCTTCGACTGTGTGATGTGGGTGCGGTACCGAACACGACGTGCGCTCGG 252

Db 197 CGGGCGACCGGCTCTCTGACGCTGCGGTGCGGCGGCGCCATGCGCATCGCC 256

QY 253 CGCGAACCGGATCCAGATCACCGGATCACCGCTCAGCCAGGTGACAGTGGCCATGCC 312

Db 257 GCGACTTCGGGATCCAAAGTGTGTCGATCACCATCAACAGAGTACAGGTGAACCGCGCC 316

QY 313 GCTATTGCGGACCGGAAACGGGACTAAGCCACCGGGT---GACTTCTGTGTGTGAT 369

Db 317 CGCGCGCAACACCGAAGGCGGCTCGACTCCCGGCGTGGAGGTGTGTGGGGAAC 376

QY 370 GCCATGTCCTCGCGCTACCGGACCAATGCTTTTCGACCGCGCTGGCCATGACATCGCTG 429

Db 377 TTCCTTCATGCTCCCTTCGCGAGCGCTCTTCGACGCGCGCTACTCATCGAGCCACC 436

QY 430 TTGAGATGTCGGAACCGGACCGTCCATCCGGAATCCTTGAGTACTCAACCGCGT 489

Db 437 TGCCACCGCGCCAGAGCTGCAAGAGCTACGCGGCGCAAGTCAACGCTGCTCAAGCCGCGC 496

QY 490 GGCATCTCGGCGCTCACCGAGGTGTCGA 517

Db 497 GGGCTTACGTTTCTACGAGTGGGTCA 524

RESULT 13

LOCUS CD442528 666 bp mRNA linear EST 03-JUN-2003

DEFINITION E101N0412G05.b Endosperm\_4 Zea mays cDNA, mRNA sequence.

ACCESSION CD442528

VERSION CD442528.1 GI:31358171

KEYWORDS EST.

SOURCE Zea mays

ORGANISM Zea mays

REFERENCE 1 (bases 1 to 666)

AUTHORS Lai,J., Dey,N., Kim,C.S., Bharti,A.K., Rudd,S., Mayer,K.F., Larkins,B., Becraft,P. and Messing,J.

TITLE Characterization of the maize endosperm transcriptome and its comparison to the rice genome

JOURNAL Genome Res. 14 (10), 1932-1937 (2004)

COMMENT Contact: lai, jinsheng

Dr. Joachim Messing's lab

Waksman Institute, Rutgers University

190 Frelinghuysen Rd., Piscataway, NJ 08854, USA

Tel: 732-445-3801

Fax: 732-445-5735

Email: jlai@waksman.rutgers.edu

Seq primer: T3.

Location/Qualifiers

1. 666

/organism="Zea mays"

/mol\_type="mRNA"

/cultivar="W22"

/db\_xref="taxon:4577"

/tissue\_type="Endosperm\_4"

/clone\_1lb="Endosperm\_4"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

ORIGIN

Query Match 9.4%; Score 77.6; DB 6; Length 666;

Best Local Similarity 51.8%; Pred. No. 4.6e-08;

Matches 201; Conservative 0; Mismatches 184; Indels 3; Gaps 1;

QY 133 TCTTGCGACAGCGCGCGCGCTCACCGACTTGTGCGGAAACGACCGTGTGAT 192

Db 198 TCCACACCGGACGCGCACGCGCTCCACGAGAGCGCTGCGGACCTCTCGGCCCGC 257

QY 193 GCGGCGCTTCGACTGTGTGATGTGGGTGCGGTACCGAACACGACGTGCGCTCGG 252

Db 258 CGGGCGACCGGCTCTCTGAGTGTGGGTGCGGCGTCCGCGGCGCCATGCGCATCGCC 317

QY 253 CGCGAACCGGATCCAGATCACCGGATCACCGTTCAGCCAGGTGCAAGTGGCCATGCC 312

Db 318 GCGCACTCGGGGTCGACGTCGTGGGATCACCATCAAGAGTACCAGGTAAGCGCGCC 377  
QY 313 GCTGATTGGCAGCGGACGCGACTAAG---CCACCGGGTGAATCTTCGTGCGTCGAT 369  
Db 378 CCGGCGCAACCGGACGAGGCGCGCTCGACTCCCCCGTCGACGAGTGTCTGCGGCAAC 437  
QY 370 GCCATGTCCTGCGATGACCGGACCAATGCTTTGACGCGCCGTTGGGCAATGACGTGCTG 429  
Db 438 TTCTCTCATGCTCTTCCCGAGCGGCTCTTGACGGCGCCTACTCTCATCGAGGCACC 497  
QY 430 TTGAGAGATCTCGAACCGGACCGGACCGTCATCCGGAATCTTTCAGTACTCAACCCGGT 489  
Db 498 TCGCAGCGCGCCAGGCTGACGAGACGTCACGGCAGGTCTACCGCGTCAAGCGCGGG 557  
QY 490 GGCATCTCGGCGCTCACCGAGTCTCA 517  
Db 558 GCGCTCTACGTCCTCTCAAGAGTGGTCA 585

RESULT 14  
CC717854 874 bp DNA linear GSS 19-JUN-2003  
LOCUS OGTAK10TM.ZM.0.7.1.5.KB.Zea.mays.genomic.clone.ZMMBma0389B19,  
DEFINITION genomic survey sequence.  
ACCESSION CC717854  
VERSION CC717854.1 GI:32122630  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 874)  
WhiteIaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,  
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
Consortium for Maize Genomics  
Unpublished (2002)  
Other\_GSSs: OGTAK10TC  
Contact: Cathy WhiteIaw  
TIGR

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whiteIaw@tigr.org  
Seq primer: TR  
Class: sheared ends.  
Location/Qualifiers  
1..874  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBma0389B19"  
/note="Vector: pBCSK-; Site\_1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

FEATURES  
source

ORIGIN

Query Match 9.4%; Score 77.6; DB 9; Length 874;  
Best Local Similarity 51.8%; Pred. No. 4.7e-08;  
Matches 201; Conservative 0; Mismatches 184; Indels 3; Gaps 1;  
QY 133 TCCTGACAGAGCGCGCGACCGGCTGACCGACCTTGTGCGGAACGAGACCGTCTGAT 192  
Db 452 TCCACCGGACGACGCGGCTGACGAGAGGCGGTGCGACCTCTCGGCGCGCG 511  
QY 193 GCGGCGCTTGAATGCTGATGTGGGTGCGTAACGGAACAACAGCGCTGCGGTGCG 252  
Db 512 CCGGCGCACCGCTCTCTGACGTGGGCTGCGGCTCGCGCGGCGCCATGCGCATGCGC 571  
QY 253 CGGACAAACGAGATCAGATCACCAGCATCAGTCAAGGTCGCAAGTGGCATGCGC 312

Db 572 GCGCACTCGGGGTCGACGTCGTGCGCATCACCATCAACGAGTACCAGGTGAAGCGCGCC 631  
QY 313 GCTGATTGGCAGCGGACGCGACTAAG---CAACCGGGTGAATCTTCGTGCGTCGAT 369  
Db 632 CCGGCGCAACCGGACGAGGCGCGCTCGACTCCCCCGTGGAGAGTGTCTGCGGCAAC 691  
QY 370 GCCATGTCCTGCGATGACCGGACCAATGCTTTGACGCGCCGTTGGGCAATGACGTGCTG 429  
Db 692 TTCTCTCATGCTCTTCCCGAGCGGCTCTTGACGGGCGCCTACTCTCATCGAGGCACC 751  
QY 430 TTGAGAGATCTCGAACCGGACCGGACCGTCATCCGGAATCTTTCAGTACTCAACCCGGT 489  
Db 752 TCGCAGCGCGCCAGGCTGACGAGACGTCACGGCAGGTCTACCGCGTCTCAAGCGCGGG 811  
QY 490 GGCATCTCGGCGCTCACCGAGTCTCA 517  
Db 812 GCGCTCTACGTCCTCTCAAGAGTGGTCA 839

RESULT 15  
CC717850/c 908 bp DNA linear GSS 19-JUN-2003  
LOCUS OGTAK10TC.ZM.0.7.1.5.KB.Zea.mays.genomic.clone.ZMMBma0389B19,  
DEFINITION genomic survey sequence.  
ACCESSION CC717850  
VERSION CC717850.1 GI:32122626  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 908)  
WhiteIaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,  
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
Consortium for Maize Genomics  
Unpublished (2002)  
Other\_GSSs: OGTAK10TM  
Contact: Cathy WhiteIaw  
TIGR

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whiteIaw@tigr.org  
Seq primer: TR  
Class: sheared ends.  
Location/Qualifiers  
1..908  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBma0389B19"  
/note="Vector: pBCSK-; Site\_1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

FEATURES  
source

ORIGIN

Query Match 9.4%; Score 77.6; DB 9; Length 908;  
Best Local Similarity 51.8%; Pred. No. 4.7e-08;  
Matches 201; Conservative 0; Mismatches 184; Indels 3; Gaps 1;  
QY 133 TCCTGACAGAGCGCGCGACCGGCTGACCGACCTTGTGCGGAACGAGACCGTCTGAT 192  
Db 738 TCCACCGGACGACGCGGCTGACGAGAGGCGGTGCGACCTCTCGGCGCGCG 679  
QY 193 GCGGCGCTTGAATGCTGATGTGGGTGCGTAACGGAACAACAGCGCTGCGGTGCG 252  
Db 678 CCGGCGCACCGCTCTCTGACGTGGGCTGCGGCTGCGGCGGCGCCATGCGCATGCGC 619  
QY 253 CGGACAAACGAGATCAGATCACCAGCATCAGTCAAGGTCGCAAGTGGCATGCGC 312

```
Db      618 GGGCACTGGGGGTCCGACGTCGTCCGCATCACCATCAACGAGTACAGGTGAAGCGGCC 559
Oy      313 GCTGATTGGGCACGCGAACCAGACTTAAG---CCACCGGGTGGACTTCTCGTGGTCGAT 369
Db      558 CGCGCGCACAAACCGCAAGGCCCGCCTCGACTCCCGCCGCTGGAGAGTCGTGCGGCAAC 499
Oy      370 GCCATGTCCTCGCCGCTACCCGGACAAATGCTTTGACGCGCCGCTGGGCCATGCAATCGCTG 429
Db      498 TTCTCTCCATGCTCCCTTCCCGACGCGATCTTGAAGGGCGCTTCACTCGAAGGCCAAC 439
Oy      430 TTGGAGATGTCGGAACCGGACCGTCATCCGGGAATCCTTCGAGTACTCAAAACCGGT 489
Db      438 TCCCAAGCGGCTCAGGAGCGTCTACGGGAGGTCTTACCGCGTCTCAAGCCGGGG 379
Oy      490 GGCATCTTCGGCGTCAACGAGGTGCTCA 517
Db      378 CGCCTCTAGTCTCTCTACGAGTGGGTCA 351
```

Search completed: August 29, 2005, 17:07:58  
Job time : 3425 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: August 29, 2005, 14:56:14 ; Search time 699 Seconds  
(without alignments)  
7750.613 Million cell updates/sec

Title: US-10-069-353A-7

Perfect score: 828 1 gtcgtgcgcgcgtgcgcacc.....cgcggaagccgcgtcgtga 828

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7331713 seqs, 3271544945 residues

Total number of hits satisfying chosen parameters: 14663426

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

1: /cgn2\_6/ptodaca/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodaca/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodaca/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodaca/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodaca/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodaca/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodaca/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodaca/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodaca/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodaca/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodaca/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodaca/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodaca/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodaca/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodaca/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodaca/1/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodaca/1/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodaca/1/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodaca/1/pubpna/US10G\_PUBCOMB.seq:\*  
20: /cgn2\_6/ptodaca/1/pubpna/US10H\_PUBCOMB.seq:\*  
21: /cgn2\_6/ptodaca/1/pubpna/US10I\_PUBCOMB.seq:\*  
22: /cgn2\_6/ptodaca/1/pubpna/US10J\_NEW\_PUB.seq:\*  
23: /cgn2\_6/ptodaca/1/pubpna/US11\_PUBCOMB.seq:\*  
24: /cgn2\_6/ptodaca/1/pubpna/US11A\_PUBCOMB.seq:\*  
25: /cgn2\_6/ptodaca/1/pubpna/US60\_NEW\_PUB.seq:\*  
26: /cgn2\_6/ptodaca/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	828	100.0	80161	17	US-10-329-148A-1
2	170	20.5	846	16	US-10-132-134-21
3	170	20.5	52101	16	US-10-132-134-1
4	165.4	20.0	80161	17	US-10-329-148A-1
5	141	17.0	88400	21	US-10-844-716-1
6	137.6	16.6	849	15	US-10-156-761-925
7	137.6	16.6	9025608	15	US-10-156-761-1

Result No.	Score	Query Match	Length	DB ID	Description
8	133.2	16.1	852	10	US-09-953-348-36
9	133.2	16.1	852	15	US-10-267-255-36
10	133.2	16.1	53500	10	US-09-953-348-76
11	133.2	16.1	53500	15	US-10-267-255-76
12	120.2	14.5	30000	11	US-09-980-217-1
13	93.8	11.3	1773	19	US-10-437-963-72340
14	93.2	11.3	1092	17	US-10-259-194A-251
15	89.4	10.8	1418	17	US-10-259-194A-529
16	84	10.1	1217	17	US-10-369-493-72721
17	83.2	10.0	828	10	US-09-953-348-35
18	83.2	10.0	828	15	US-10-267-255-35
19	76	9.2	1733	20	US-10-425-115-130152
20	72.4	8.7	942	10	US-09-942-025-8
21	72.4	8.7	67311	10	US-09-942-025-1
22	71.2	8.6	795	17	US-10-369-493-35691
23	69.8	8.4	1291	18	US-10-425-114-14785
24	69.8	8.4	1393	18	US-10-424-599-27555
25	66.6	8.0	741	19	US-10-767-701-20510
26	63.2	7.6	1291	18	US-10-425-114-7357
27	63.2	7.6	1528	18	US-10-424-599-27554
28	59	7.1	1862	16	US-10-149-759-79
29	58.4	7.1	2757	17	US-10-282-122A-26027
30	58.2	7.0	1287	15	US-10-156-761-3515
31	57.2	6.9	1152	9	US-09-801-368-95
32	57.2	6.9	1152	17	US-10-369-493-25549
33	57.2	6.9	1152	19	US-10-793-639-523
34	57.2	6.9	1320	9	US-09-779-144A-1
35	57.2	6.9	1320	9	US-09-779-144A-5
36	54.6	6.6	612	17	US-10-369-493-45526
37	52.8	6.4	1230	19	US-10-437-963-65029
38	52.6	6.4	858	15	US-10-156-761-871
39	52.6	6.4	9025608	15	US-10-156-761-1
40	52.4	6.3	2560	19	US-10-437-963-3228
41	52.2	6.3	738	15	US-10-156-761-216
42	52	6.3	783	19	US-10-767-701-2990
43	52	6.3	819	15	US-10-156-761-1869
44	52	6.3	2004	17	US-10-282-122A-28310
45	51.6	6.2	624	17	US-10-369-493-44312

# ALIGNMENTS

RESULT 1  
US-10-329-148A-1  
; Sequence 1, Appl1  
; Publication No. US20040023343A1  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H  
; APPLICANT: Broughton, Mary C  
; APPLICANT: Crawford, Kathryn P  
; APPLICANT: Madhuri, Krishnamurthy  
; APPLICANT: Treadway, Patti J  
; APPLICANT: Turner, Jan R  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; FILE REFERENCE: 50489 Div1  
; CURRENT APPLICATION NUMBER: US/10/329,148A  
; PRIOR FILING DATE: 2002-12-23  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/603,207B  
; PRIOR FILING DATE: EARLIER FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/370,700  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 80161  
; TYPE: DNA  
; ORGANISM: Saccharopolyspora spinosa  
US-10-329-148A-1  
Query Match 100.0%; Score 828; DB 17; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 6.5e-224;

[illegible]

```

; SOFTWARE: Patentin version 3.0
; SEQ ID NO 21
; LENGTH: 846
; TYPE: DNA
; ORGANISM: Streptomyces platensis subsp. rosaceus
US-10-132-134-21

```

Query Match	20.5%	Score 170;	DB 16;	Length 846;
Best Local Similarity	59.4%	Pred. No. 4.3e-38;		
Matches 328; Conservative	0;	Mismatches 215;	Indels 9;	Gaps 2;

OY 5 TGGCAGGTGGCGCACCAACATGCAATGCGAGCGAGTTGGCGAGATGTATGACCTGTACCGCGGT 64  
 Db 35 TCCCTCTCTCCGCCCCCTCGTCCCGGAAGAGTGGACACTTATGACCGCGCTCACCGCAC 94  
 OY 65 TGTGTAATCGATCGGTGGCGGCGGCCCTCGGCATTCACACAGCGTACTGGAGAGACGACG 124  
 Db 95 TGGACACCGGAAGCGCGCGCG-----GAGGCTTCACCTTGCTTACGGGAGCGTGAAG 148  
 OY 125 GCGGCGGCTTCTGGC---AGCAGGCGCGCGACCGGCTCACCGACTTGTGGCGGAACGGA 181  
 Db 149 ACAACGACACCCCGGCTCGTGAAGCGCGCGACCGGCTCACCGACGATGACGACCGCC 208  
 OY 182 CCGTGTCGATAGCGCGCGCGTTCGACTGCTCGATATGTGGGTGGGATACCGGACACACGCGC 241  
 Db 209 TGGGATATGACCAAGGACAGCGGGATCTTCGACAGTGGGTGGAGTGGGCGACGCGGCA 268  
 OY 242 TGGCGATCGCGCGCGCGACAACGCGATTCAGATTCACCGGATCACCGTACGCGAGGTGCAG 301  
 Db 269 TGGGATATGCGCGCGCGCGACCGGCGCGCCATGTGTCAGGGGATGGCGATACGCAAGGACAGA 328  
 OY 302 TGGCCATCGGCGGCTGATTGCGCAGCGCAACGCGGACTTAAGCCACCGGAGTGAATTCTCGT 361  
 Db 329 TCGCCGCGCGCACCGGCGCTCGCGAGGCGCGCGGCTGAGCGACCGGTTGAGTTCCGCGC 388  
 OY 362 GCGTGCATGCAATGTGCTCCGCGCGCTACCGCGACCAATGCTTTCGACGCGCGCTCGGSCATGC 421  
 Db 389 AGCGCGACGCGATGGAATGCTCCCTTCCCGACAGACTCTTTCGACGCGCGCATGGCCATTCG 448  
 OY 422 AGTGCCTGTTGGAGATGTCCGAACCGGACCGGTGCATTCGGGAATCTTTCGATGTACTCA 481  
 Db 449 AGTGCATCTTTCACATGAGCCCGACCGCGGACCGGATCTGCGGAAATCCGCGCGGTACTGC 508  
 OY 482 AACCCGATGACATCTCGGCGGTACACGAGGTGTCAACGAAAGCGGAGCGCGGATGC 541  
 Db 509 GCCCGGAGCGCGCGCTGATCTCAACGACTTCTTCGAGCGCGCGCCGTCCCGCGCAGA 568  
 OY 542 CGGTATCCGGG 553  
 Db 569 AGCAGGCCCGG 580

```

RESULT 3
US-10-132-134-1
Sequence 1, Application US/10112134
Publication NO. US2003071562A1
GENERAL INFORMATION:
APPLICANT: Farnet, Chris
APPLICANT: Yang, Xianhu
APPLICANT: Staffa, Alfredo
APPLICANT: Zazopoulos, Emmanuel
TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
FILE REFERENCE: 3013-2US
CURRENT APPLICATION NUMBER: US/10/132,134
CURRENT FILING DATE: 2002-04-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: Patentin version 3.0
SEQ ID NO 1
LENGTH: 52101
TYPE: DNA
ORGANISM: Streptomyces platensis subsp. roseaceus
US-10-132-134-1

```



Query Match 20.5%; Score 170; DB 16; Length 52101;  
Best Local Similarity 59.4%; Pred. No. 5.1e-38;  
Matches 328; Conservative 0; Mismatches 215; Indels 9; Gaps 2;

```
QY 5 TGGCAGTGGCGGACCAACATCGACAGAGTTGGGAGATGTATGACTGTGATCAAGCCGT 64
DB 51268 TCCCCCTCTCCGCCCCCGTCGCCGAGAGGTGGACACCTCTACGACCCGCTACCGGCAC 51327
QY 65 TGTGAACTCGGTGCGGGGCGCCCTGCGCATTCACACCGGCTACTGGAGAACAGAC 124
DB 51328 TGGACACCGAAGGCGCGG-----GACGCTTCCACCTCGGCTACTGGAGAGTGGAGC 51381
QY 125 GGGCGGCTTCTGCG---AGCAGCGCGCGGACCGGCTCAACCACTTGTGCGCGAACGA 181
DB 51382 ACNACGACACCCGCTGTGGAGCGGCCGACCGGCTCAACGATGACCGACCGCC 51441
QY 182 CGGTGTCATGAGCGCGGCTTTCGACTGTGATGGGGGTGCGGTACCGGACCAAGCGC 241
DB 51442 TGGGATTCGACGAGGACAGCGGCTCTTCGACGTGGCTGGAGTCCGCGACCGGCCA 51501
QY 242 TGGCGCTGCGCGCGACCAACGCGATCCAGATCAACCGCATCACCGTCAAGCAGTGCAG 301
DB 51502 TGGGATTCGCGCGCGGACCGGCGCATGTCAACCGGATCCGATCAGCAGACGACGA 51561
QY 302 TGGCATTCGCGCGCTGATGGGACGCGGACCGGACTAAGCCACCGGTTGGACTTCTCT 361
DB 51562 TCGCCCGCGCGACCGGCTTTCGCGAGGCGCGCGCTGACGACCGCGTGGAGTTCCGCG 51621
QY 362 GCGTGCATGCGATGCTCCGCTGCGCGGACCGGACATGCTTTCGACGCGCGCTGGGACATG 421
DB 51622 ACCCGACGCGCATGAGACTGCTTCCCGACGATCTTTCGACGCGCGCATGCGCATG 51681
QY 422 AGTGTGCTTGGAGATGTCCGAACCGGACCGGTCCATCCGGAATCTTTCAGATCTCA 481
DB 51682 AGTGTGCTTTCGACATGCGCGGACCGGATCTTTCGCGAGATCCGCGGCTACTG 51741
QY 482 AACCCGCTGCGATCTTTCGCGGCTCACCGAGGTCTTCGACGAGAACGAGCGGCGGATG 541
DB 51742 GCCCGCGCGCGCGCTGTGTCTTTCGACGATCTTTCGAGCGGCGCGCGCGGAGA 51801
QY 542 CGGTGTCCGCGG 553
DB 51802 AGCAGCTCCGCG 51813
```

## RESULT 4

US-10-329-148A-1/c  
Sequence 1, Application US/10329148A  
Publication No. US20040023343A1  
GENERAL INFORMATION:  
APPLICANT: Baltz, Richard H  
APPLICANT: Broughton, Mary C  
APPLICANT: Crawford, Kathryn P  
APPLICANT: Madhuri, Krishnamurthy  
APPLICANT: Treadway, Patti J  
APPLICANT: Turner, Jan R  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Bio-synthetic Genes For Spinosyn Insecticide  
FILE REFERENCE: 50489 DIV1  
CURRENT APPLICATION NUMBER: US/10/329, 148A  
CURRENT FILING DATE: 2002-12-23  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/603, 207B  
PRIOR FILING DATE: EARLIER FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/370,700  
PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-09  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 80161  
TYPE: DNA  
ORGANISM: Saccharopolyspora spinosa  
US-10-329-148A-1

Query Match 20.0%; Score 165.4; DB 17; Length 80161;  
Best Local Similarity 52.8%; Pred. No. 1e-36;  
Matches 386; Conservative 0; Mismatches 336; Indels 9; Gaps 1;

```
QY 15 CCGACCAACATCGACAGAGTTGGGAGATGTATGACTGTGACCGGCTGTAATC 74
DB 13575 CACACAGGATTTCCGACCGCGGATCAGGTGAGTCAATCTTCATGCTGTCGCGCAC 13516
QY 75 GGTTCGGGCGGCGCCCTGCGCATTCACACCGGCTACTGGAGAACAGACGGCGGCTTC 134
DB 13515 GGTTCCTTCGACACAGGTTTACGAGCGGCGGATTCGAGAGATGTCGTCGACACC 13456
QY 135 CTGGACAGAGGCGCGGACCGGCTCACCGACTTGTGCGGACGACCGGCTCGATG 194
DB 13455 GTGTGCGAGTGTGCGGACCACTGACGACTTGTTCATGACAGAGCGGCGCTCCGTTC 13396
QY 195 CCGGCTTCAGTCTGATGTGGGTGCGGTACCGGACCAACGAGCGCTGCGCGCG 254
DB 13395 CCGAGCGACCTGTTGACCTGGGCTGGGCAATGGGACGCGGATGTCGCGCATG 13336
QY 255 CCGACACCGGATTCAGATCAGGATCAGCGGATCACCGTCAAGCGAGTGCAGTGGCCATGCGCG 314
DB 13335 CCGCAGCGGCGTTCGAGTCAAGTCAAGGATCACGTTGAACCGCCAGCATCTGCGCGCAC 13276
QY 315 TGATTGGCAGCGGAAACGCGGACTAAGCCAGCGGCTGACTTCTGTCGTCGATGCAAT 374
DB 13275 CAGGCTTCGCAACGAGACCGGACTGCGCGGCTGTTGATGATCTGATGTCAGCGCG 13216
QY 375 GTCCCTGCGGATCCCGGACATGCTTTCGACGCGCGCTGGGCGCATGCTGCTTTCGA 434
DB 13215 CAGGCTGCGGATCCCGGACGATGCTTTCGACGCGCGGCTGATGATGCTGCTGCA 13156
QY 435 GATGTCCGACCGGACCGGCTGATCAGCGGAAATCTTTCGAGTCAACTCAACCGGTCGAT 494
DB 13155 GATGTGACGACGCGCGCGGATTCGCGGATTCGCGGATTCGAGAACCGGCGCGCG 13096
QY 495 CCGTGGCGTTCACCGGAGTGTCAACGAGAGCGGCGGCGGATCCGAGTGTCCGCGGA 554
DB 13095 GTTGTGCTTCGAGATCATCTGCGGTTGATCTCCGGAAGATGACCGCGGCTTTCG 13036
QY 555 CAGGTGCGGACCGGCTTTCGATGCTGCTGCTGAGCACTTTCGAAATGCTGCTGCTG 614
DB 13035 GACGGGACAGACCGGCTCATACCTT-----GACGACTTCACGCGCGCTGGTACGCA 12985
QY 615 AGCGGAGTTTCGAGATCTTCGATTTGGAGACGTTGTTCGAGGACCGGCTGATTCATGCG 674
DB 12984 AGCGGAGTTTCGAGATCTTCGAGTTCACGACCTTCAGCGGACAGACGAGTGCATGCTTC 12925
QY 675 GCAGTTTCGCGGAGAGCTGCTGCGGACGAGGATTCGCGGATTCGCGGATTCGCGGCG 734
DB 12924 CTGTAGTGTGACGAGATTCCTCGGAACTCGATGAGCTTCGCGGCTGCGGCTGCGG 12865
QY 735 TGTGCGCGGCT 745
DB 12864 TGTGCGGACT 12854
```

## RESULT 5

US-10-844-716-1/c  
Sequence 1, Application US/10844716  
Publication No. US20050003409A1  
GENERAL INFORMATION:  
APPLICANT: Huang, Chengjin  
APPLICANT: Chaleff, Deborah T.  
APPLICANT: Ruppen, Mark E.  
APPLICANT: Stephens, Jerome  
TITLE OF INVENTION: Cloning Genes from Streptomyces Cyaneogriseus Subsp.  
TITLE OF INVENTION: Noncyanogenous Genes for Biosynthesis of Antibiotics and Methods of Use  
FILE REFERENCE: AM100484  
CURRENT APPLICATION NUMBER: US/10/844, 716  
CURRENT FILING DATE: 2004-05-13  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.2

```

; SEQ ID NO 1
; LENGTH: 88400
; TYPE: DNA
; ORGANISM: bacteria
US-10-844-716-1

```

Query Match	17.0%;	Score 141;	DB 21;	Length 88400;
Best Local Similarity	56.4%;	Pred. No. 8.1e-30;		
Matches 286;	Conservative	0;	Mismatches 215;	Indels 6;
				Gaps 1;

Qy	13	GGGCGACCAACATGTGAGCAGAGGTTGGGCACATGATGAGCCTGGTCAAGCGCGTTGCGTGAAC	72
Db	51282	GGATCACCGAACCCCGCGAGAGCTGGGATCTACTAGACCGCATGACCAAGCTGTTGAAC	512223
Qy	73	TCGGTGGCGGAGCGGCCCTTGCGGCATTCACACAGCGCTACTGGAGAACGACGGGCGGAGCT	132
Db	51222	CGGGCCTCTGGGCGGCAACAGCACTTCGGATATCGGCGCGCACCGGACACCGGACAGCA	511633
Qy	133	TCCTGGAGAGAGGCGCGCGACCGGCTCACCGACCTTGTGCGCGAAGAACCGGTCTCAT	192
Db	51162	CTGGCCAGGCGCTTCGACCGGCTTCACGACCATATATATGGCAAGCTTGGGAAACACACC	511033
Qy	193	GG-----CGGCGTTCCACTGCTCGATGTGGGGGTGGCGTACCGAACACAGCGCTGCGCG	246
Db	51102	GGGGCGCCGGGTCAAGAGTGATGTGAGAGTGTGGCTGGGGTTCCGGCGCGCGCGGCGCTGGCGC	510433
Qy	247	GTGGCGGGCGCAACACGGGATTCAGATCACCGGATACCGTCAACGACGAGGTGCAATGGGCC	306
Db	51042	CTGGCGGACAGCGAGCGCGGTGAGCATGTGGGATACCATCAATGACGCCCGGACAGGTTCGAG	509833
Qy	307	ATGCGCGCTATTTGGCGACCGCAACCGGACCGGACTTAAGCCACCGGATGGACTTCTCGTGCATC	366
Db	50982	CTGGCGACCGCGCTCGCGCGAGGGGTCCGAGCTGGCGAACCGGGTCCGCTTTCAGATGGCGCC	509233
Qy	367	GATGCCATGTTCCTCGCCGCTAACCCGACATGCTTTTGAAGCGCGCGCTGGGCGCATGCATGTG	426
Db	50922	GAGCGCATGGAACCTTGCGCTTTCGCCAGCGCTCTCTTTGAAGCGCGCTCGGGCGCTTCGAGTGC	508633
Qy	427	CTGTTGGAGATGTCCGAACCGGACCGTGCATCCGGGAAATCCTTTGAATGATCTCAAAAGCC	486
Db	50862	CTGCTGCACATGCGCCGACCCGGCGCGGGATTTCCAGAGATAGGCCCGGTGTGCTGCGCGCCC	508033
Qy	487	GGTGGCATCTCGGCGCTCACCGAGTTC	513
Db	50802	GGTGGGCGGCTGGCGCGCATGAGAGTTC	50776

## RESULT 6

```

US/10-156-761-925
# Sequence 925, Application US/10156761
# Publication No. US20030119018A1
# GENERAL INFORMATION:
# APPLICANT: OMURA, SATOSHI
# APPLICANT: IKEDA, HARUO
# APPLICANT: ISHIKAWA, JUN
# APPLICANT: HORIKAWA, HIROSHI
# APPLICANT: SHIBA, TADAYOSHI
# APPLICANT: SAKIBI, YOSHIYUKI
# APPLICANT: HATTORI, MASAHIRA
# TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
# FILE REFERENCE: 249-262
# CURRENT APPLICATION NUMBER: US/10/156,761
# CURRENT FILING DATE: 2002-05-29
# PRIOR APPLICATION NUMBER: JP 2001-204089
# PRIOR FILING DATE: 2001-05-30
# PRIOR APPLICATION NUMBER: JP 2001-272697
# PRIOR FILING DATE: 2001-08-02
# NUMBER OF SEQ ID NOS: 15109
# SEQ ID NO 925
# LENGTH: 849
# TYPE: DNA
# ORGANISM: Streptomyces avermitilis
# FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: (1) .. (849)
US-10-156-761-925

```

Query Match	16.6%	Score 137.6;	DB 15;	Length 849;
Best Local Similarity	50.0%	Pred. No. 6.2e-29;		
Matches 372; Conservative	0;	Mismatches 369;	Indels 3;	Gaps 1;

Qy	15	CGACACCAATCGCAGCAGGTGGGCAATGATGACCTGGTCAAGCCGTTCGAACTC	74
Db	24	CCGACCTCTTCCCTCGAAGGTGGGGACCTACTACACACGTTGTGACCGAATCTAGAAATCG	83
Qy	75	GGTGCGGGCGGCGCCCTGGCCCATCACACAGGCTACTGGAGACAGCGGGCGGCTTC	134
Db	84	TGGCGTGGGTGGGAACACCACTCGAATCTGGCGCGAGCGGGGGACGGCAG---TTC	140
Qy	135	CTGGCAGCAGGCGCGCGACCGGCTCACCGACCTTGTGCGCGAACGACCGTGTCTGATGG	194
Db	141	ACCGGCGAAGGCGCGCGACCGGCTCACCGAATCTCTTCATCGGCAAGCTGAGAGGCATCAC	200
Qy	195	CGGCGTTGCACTGCTGATGTGTGGGTGGGGTATCCGACACACAGCGCTGGCGGCTGGCGG	254
Db	201	GGGCGCGCGGGGTCTGGAGCTCGGGCTGGGTTTCGGAAAGCGCGGGGTGGCGCTGGCTT	260
Qy	255	CGACACGGGATCTCAGATCACCGGATCACCGTCAAGCAGGATGGACATCGCGC	314
Db	261	GAGCGGCGCGCTGATGTGTGTGGGTACGGGTAGCGAGGTTTAAAGTGGGCTGGCGAC	320
Qy	315	TGATTCGCGACGCGAACCGGACCTAAGCCACCGGGTGGACTTCTGTGCTGATGTCAT	374
Db	321	CGCTTCGCGAAGCAGTGCACGTGGCGGACCGGGGTGTGTTTCAACCGTGCAGCGCAT	380
Qy	375	GTCCCTCCCGTACCCGGAACAATGCTTTTGAACGCCGCTGGGCCATGCAATGCTGTTTGA	434
Db	381	GGAACTGCGGTTTCCGACAGGGTCTTTGACGCGCGGTGGCCCTTGAAGTGTCTCTGCA	440
Qy	435	GATGTCGAACCGGACCGTGCATCCGGGAATCTTTCAGATCTCAAAACCGGTGGCAT	494
Db	441	CATGCCCAAGCCCGCACAGTGAATCCGGGAATATGCCCGGGTGTCTCGCCCTCGGCGCG	500
Qy	495	CTTCGGCGTCAACGAGTGTCTCAAAAGAAACGAGCGGGCGCGGATGCGGTTCTCGGGA	554
Db	501	GCTGGCGGTCAAGGACGTGCACTTGCAGCGCTTGGGGGAGCCGGCATGAAAGCGCGGGGA	560
Qy	555	CAGGTGGCGGACCGGCTTCGGAATTCGTGCTGGCTGAGCACTTTTGGAAATCGCTGCTGC	614
Db	561	GTCACATGCCCAAGTTGCTTCGCGGTTCGCGCTCGGCTCTGTGTGACATTCACACAGATCCCGGAAT	620
Qy	615	AGCGGGATTTCAGATCTTCGATTGGAGAGAGGTGTGTCGAGACCGGGATCTTCAATGCC	674
Db	621	GATCGCGCAAGCGCGGGCTGGAACTGCAATGAACTGACCGACATTCGGCGATCAGATCTGTGG	680
Qy	675	GCAGTTTCGCCGAAGACTTCGCTGCGCACACGAGATTCGCGACAGTACGGGCGCGC	734
Db	681	CCCTCTTTTCGCGGCTGCGTGACCAAGTAAAGACACCTGCAGCAGATACCGGGGCG	740
Qy	735	TGTTCGCGGCTGGGCGCGCGGT	758
Db	741	CTTTCGGGATCGCGTTCGCGAAT	764

## RESULT 7

US-10-156-761-1/c  
Sequence 1, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMIKA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA

```

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

```

Query Match      16.6%; Score 137.6; DB 15; Length 9025608;
Best Local Similarity 50.0%; Pred. No. 8.8e-29;
Matches 372; Conservative 0; Mismatches 369; Indels 3; Gaps 1;

```

```

QY 15 CGACCAACATCGCAGAGGTGGCAGATGATGACCTGTGACCGCGTGTGCTGAATC 74
DB 1137457 CCGCCCTCTTCTCGAAGTGGAGGACTAATGACCGTTTGAACGACTTCAATGATCG 1137398
QY 75 GGTGCGGCGCGCGCCCTGCGCATCCACACGCGCTACTGGAGAACGACGCGCGGCTTC 134
DB 1137397 TGGCGTGGGTGGAGAACCCACCTCGGATCTGCGCGACGCGGGGAGCGGAG---TTC 1137341
QY 135 CTGGCAGACGCGCGCGCGCTCACCGACTTGTGCGCGAAGCAGCCTGCTCGATGG 194
DB 1137340 ACCGGGCAAGCGCGCGACCGGCTCACCGATCTCTATCGGCAAGCTGAGAGGCATCAC 1137281
QY 195 CGCGCTTTCATCTGTGATGTGGGTGGGTACCGGACACACGCGCTGGCGGTGGCGG 254
DB 1137280 GGGCGCGCGGCTCTGACCTGCGCTGCGGTTCCGGAAGCCGCGGCTGCGCTCGCCCT 1137221
QY 255 CGACCAACGCGATCCAGATCAACCGGATCACCGTACGCGAGTGCAGATGCGCATTCGCGC 314
DB 1137220 GAGCGCGCGCGGTGATGTCCTGCGCGTGAAGGAGTTCAGGTTCGCGGCTGGGAGC 1137161
QY 315 TGATGGCGACGCGGACCGGACCTAAGCCACCGGCTGGAATTTCTGCTGCTGATGCTCAT 374
DB 1137160 CGCTCTCGCAAGCAGTGCAGCTGGCGGACCGGCTGTGTTCAACCGTCCGACGCGAT 1137101
QY 375 GTCCCTGCGGTACCCCGAACATGTTTTCAGACGCGCGCTGGCCATGAGTTCGCTTTGA 434
DB 1137100 GGAACCTGCGTTCCTCCGAGGCTCTTTCAGACGCGCGCTGGGCTGAGTCTCCCTGCA 1137041
QY 435 GATGTCCGAACCGGACCGTGCATCCGGAATTCCTTCAGATCTCAAAACCGGTGCGAT 494
DB 1137040 CATGCCAAGCTCCGCAAGTGAATCCGGAGATTCGCGGCTGTGTCGCCCGCGCGCG 1136981
QY 495 CTTGCGGTCACCGAGGTGCTCAAGCAAGGAAGCGCGCGCGGATGCGGTGTCGCGGGA 554
DB 1136980 GCTGCGCGTACCGGACCTGCACTGCGCGCTTTCGCGGACCGGCACTGAAGGCGCGGGA 1136921
QY 555 CAGGTGCGGACCGGCTTTCGATGTCCTGCTGCAACAATTCTGGAATTCGCTGCGTGC 614
DB 1136920 GTGACCTGCTCCAGTTGTCGCGGCTCCGCGCTGCTGTCACATGACAGATACCGCGAAT 1136861
QY 615 AGCGGAGTTCAGATCTTCGATTGGAGAGACGTGTCGTGAGAGACCGCGTACTTCATGCC 674
DB 1136860 GATGCGCGAGCGCGGAGTGAATGATGAGTGAACGACATGCGGATCAGAGTTCGTCGG 1136801
QY 675 GCAATTGCGGAGAGCTTCGCTGCGACACGACGCGGATGCGGACAGATACGCGCGCGC 734
DB 1136800 CCCCCTTTTGGCGCGCGCTGCTGACACAGTGAACGACACCTCGACGATACGCGCGCGC 1136741
QY 735 TGTGCGCGGCTGGCGCGCGCGGT 758

```

```

DB 1136740 CTTGCGGATCGCGGTGCGCGGAGAT 1136717

```

```

RESULT 8
US-09-953-348-36

```

```

; Sequence 36, Application US/09953348
; Publication No. US20030134398A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, David. H
; APPLICANT: Mao, Yingqiang
; APPLICANT: Varoglu, Mustafa
; APPLICANT: He, Min
; APPLICANT: Sheldon, Paul

```

```

; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER

```

```

; FILE REFERENCE: 600.530US1
; CURRENT APPLICATION NUMBER: US/09/953,348
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: PCT/US00/06394
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/266965
; PRIOR FILING DATE: 1999-03-12
; SOFTWARE: FaSTSeq for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 36
; LENGTH: 852
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-953-348-36

```

```

Query Match      16.1%; Score 133.2; DB 10; Length 852;
Best Local Similarity 54.0%; Pred. No. 1.1e-27;
Matches 225; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

```

```

QY 15 CGACCAACATCGCAGAGGTGGCAGATGATGACCTGTGACCGCGTGTGCTGAATC 74
DB 36 CTCACCGCGCTCGAGAGAGTGGCGCGCTTAACGACCGGTTACCGCGCTGGAGACCGCG 95
QY 75 GGTGCGGCGCGCGCCCTGCGCATCCACACGCGCTACTGGAGAACGACGCGCGGCTTC 134
DB 96 CTCCTCTGCGGAGAACCTGACCTTCTGCTAATGAGACT--CCCCGACAGCGAGGTGTC 152
QY 135 CTGGCAGACGCGCGCGGACCGGCTCACGACCTTGTGCGCGAAGGACCGTGTCTGATG 194
DB 153 GCTGCGCGGAGCGCACCGGCTCACCGATGATGCGCGGCTGCGATCGCGCGC 212
QY 195 CGCGCTTTCATCTGTGATGTGGGTGGGTACCGGACACGACGCTGCGGTGCGCGC 254
DB 213 CGCGCTCGCGGTCTGAGACTCGGCTGCGGCGGCGGAGACCGCGGCGTACGATCGCGCG 272
QY 255 CGACCAACGCGATCCAGATCAACCGGCTACCGTCAACGAGTGCAGATGCGCATTCGCGCG 314
DB 273 GCTCAGCGGAGCGCATGTACCGGCAATCTGTGAGACCATGAGCAGTGTCTCGGAGCAA 332
QY 315 TGATTCGCAACGCGGACCGGACCTAAGCCACCGGCTGAGACTTCTCGTGTGATGCTCAT 374
DB 333 CGCGCTGCGCGAGAGGCGCGGCTCGCCACCGGCGCGCTTTCAGCGGCGCGACCGAT 392
QY 375 GTCCCTGCGGTACCCCGAACATGCTTTCAGACGCGCGCTGGGCGATGCACTGCTTTGA 434
DB 393 GGAACCTCCCTTGAAGAGAGAGAGCTTCGACGCGCGTTCATGCGCTCGAATTCATTCGA 452
QY 435 GATGTCCGAACCGGACCGTGCATCCGGAATTCCTTCAGATCTCAAAACCGGTGCGAT 494
DB 453 CATGCCAAGCGCGGCTCAGGTGCTTTCAGAGTGGCGCGGCTGCTCGGCGCGGAGCGG 512
QY 495 CTTGCGGTCACCGAGGTGCTCAAGCAAGGAAGCGCGCGCGGATGCGGTGTCGCGGGA 554
DB 513 TCTGCTGCTACCGACTTCTTTCAGAGCGGCGCGCTTCGCGCGCGGAGGCGCGCGCG 572
QY 555 CAGGTG 560
DB 573 CCAGCG 578

```

RESULT 9  
US-10-267-255-36  
; Sequence 36, Application US/10267255  
; Publication No. US20030124689A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600,456US1  
; CURRENT APPLICATION NUMBER: US/10/267,255  
; PRIOR FILING DATE: 2002-10-09  
; PRIOR APPLICATION NUMBER: US 09/266,965  
; PRIOR FILING DATE: 1999-03-12  
; PRIOR APPLICATION NUMBER: US 08/624,447  
; PRIOR FILING DATE: 1996-08-19  
; PRIOR APPLICATION NUMBER: PCT/US94/11279  
; PRIOR FILING DATE: 1994-10-06  
; PRIOR APPLICATION NUMBER: US 08/133,963  
; PRIOR FILING DATE: 1993-10-07  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO: 36  
; LENGTH: 852  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-10-267-255-36

Query Match 16.1%; Score 133.2; DB 15; Length 852;  
Best Local Similarity 54.0%; Pred. No. 1.1e-27;  
Matches 295; Conservative 0; Mismatches 248; Indels 3; Gaps 1;  
QY 15 CGCACCAATCGGACGAGTTGGGACATGATGATCTGCTGACGCCGTTGCTAACTC 74  
DB 36 CTCACCCGCTCCGAGAGTGGGCGCGCTCTACACCGGTTCCACCGCGCTGGAGCGCG 95  
QY 75 GGTGCGGGCGGCCCTTGCGCCATCCACACGCGTACTGGGAGACGAGCGGGGCTTC 134  
DB 96 CTCCTCTGGGAGAACCTGCACTTGGCTACTGGAAGT---CCCCGACGACGAGTGGC 152  
QY 135 CTGGACGAGCGCGCCGACCGGCTGACCGACTTGTGCGCGAACGAGCCGTGCTGATGG 194  
DB 153 GCTGACGAGGCGACCGACCGGCTCACCGACATGATGGCGGCTGCGCATGCGGCGC 212  
QY 195 CGGCGTTGACTGCTGATGAGGTGGGTGCGGTACCGGACCAACGAGCGCTGCGGCGCG 254  
DB 213 CGGCTCCGCGCTGACCTGCGTGGCGCGGTGGGAGACCCGCGGCGTACGATGCGCG 272  
QY 255 CGACACGCGATCCAGATCAACCGGATCACCGTACGACGAGGTGCAAGTGGCCATGCGCGC 314  
DB 273 GCTCAGCGGAGCGCATGTCAACGGGCACTCTGGTGAAGCATGAGAGGTGTCCGGGCCAA 332  
QY 315 TGATTGGCAGCGGACCGGACCTTAAGCCACCGGGTGAATTCTTGTGCGTGTGATGCCAT 374  
DB 333 CGCGCTGACGAGGAGCGGCGGCTGCGGACCGGGCGCGCTTCAAGCGGCGGAGCGCAT 392  
QY 375 GTCCCTGCGGTAACCGGACATGCTTTGAGCGCGCGCTGGGCGCATGCGTGTGTTGA 434  
DB 393 GGAACCTCCCTTCGAGGACGAGAGCTTCAGCGCCGCTCATCGCCCTCGAATCGATCATCA 452  
QY 435 GATGTCGGAACCGGACCGTGCATCCGGGAAATCTTGAAGTACTCAAAACCGGATGCAT 494  
DB 453 CATGCCCAACCGCGCCGAGGTCTGCGCCAGGTCCGGCGGTCTGCGGCGCGGGGCGCG 512  
QY 495 CCTGGCGTCAACCGAGGTGTCTCAACGAGAAACGGGCGGCGGATGCCGTGTCCGGGGA 554  
DB 513 TCTGTGCTCAACGACTTCTTGAAGCGGCGCGCGCGCGCGCGCGCGCGCGT 572  
QY 555 CAGGTG 560  
| | | | |

DB 573 CCAGCG 578  
RESULT 10  
US-09-953-348-76/C  
; Sequence 76, Application US/09953348  
; Publication No. US20030134398A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, David, H  
; APPLICANT: Mao, Yingting  
; APPLICANT: Varoglu, Mustafa  
; APPLICANT: He, Min  
; APPLICANT: Sheldon, Paul  
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER  
; FILE REFERENCE: 600,530US1  
; CURRENT APPLICATION NUMBER: US/09/953,348  
; PRIOR FILING DATE: 2001-09-12  
; PRIOR APPLICATION NUMBER: PCT/US00/06394  
; PRIOR FILING DATE: 2000-03-10  
; PRIOR APPLICATION NUMBER: 09/266965  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 153  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 76  
; LENGTH: 53500  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-09-953-348-76

Query Match 16.1%; Score 133.2; DB 10; Length 53500;  
Best Local Similarity 54.0%; Pred. No. 1.3e-27;  
Matches 295; Conservative 0; Mismatches 248; Indels 3; Gaps 1;  
QY 15 CGCACCAATCGGACGAGTTGGGACATGATGATCTGCTGACGCCGTTGCTAACTC 74  
DB 7549 CTCACCCGCTCCGAGAGTGGGCGCGCTCTACACCGGTTCCACCGGCTGGAGCGCG 7490  
QY 75 GGTGCGGGCGGCCCTTGCGCCATCCACACGCGTACTGGGAGACGAGCGGGGCTTC 134  
DB 7489 CTCCTCTGGGAGAACCTGCACTTGGCTACTGGAAGT---CCCCGACGACGAGTGGC 7433  
QY 135 CTGGACGAGCGCGCCGACCGGCTGACCGACTTGTGCGCGAACGAGCCGTGCTGATGG 194  
DB 7432 GCTGACGAGGCGACCGACCGGCTCACCGACATGATGGCGGAGCGCTGCGCATGCGCGC 7373  
QY 195 CGGCGTTGACTGCTGATGAGGTGGGTGCGGTACCGGACCAACGAGCGTGGCGCGCGCG 254  
DB 7372 CGGCTCCGCGCTTGAAGCTCTGCGTGGCGGTGGGAGACCCGCGGCTTACGATGCGCG 7313  
QY 255 CGACACGCGATCCAGATCAACCGGATCACCGTACGACGAGGTGCAAGTGGCCATGCGCGC 314  
DB 7312 GCTCAGCGGAGCGCATGTCAACGGGCACTCTGGTGAAGCATGAGAGGTGTCCGGGCCAA 7253  
QY 315 TGATTGGCAGCGGACCGGACCTTAAGCCACCGGGTGAATTCTTGTGCGTGTGATGCCAT 374  
DB 7252 CGCGCTGACGAGGAGCGGCGGCTGCGGACCGGGCGCGCTTCAAGCGGCGGAGCGCAT 7193  
QY 375 GTCCCTGCGGTAACCGGACATGCTTTGAGCGCGCGCTGGGCGCATGCGTGTGTTGA 434  
DB 7192 GGAACCTCCCTTCGAGGACGAGAGCTTCAGCGCCGCTCATGCGCTCGAATCGATCATCA 7133  
QY 435 GATGTCGGAACCGGACCGTGCATCCGGGAAATCTTGAAGTACTCAAAACCGGATGCAT 494  
DB 7132 CATGCCCAACCGCGCCGAGGTCTGCGCCAGGTCCGGGTGTGCGGCGCGGGGCGCG 7073  
QY 495 CCTGGCGTCAACCGAGGTGTCTCAACGAGAAACGGGCGGCGGATGCCGTGTCCGGGGA 554  
DB 7072 TCTGTGCTCAACGACTTCTTGAAGCGGCGCGCGCGCGCGCGCGCGCGT 7013  
QY 555 CAGGTG 560  
DB 7012 CCAGCG 7007  
| | | | |

RESULT 11  
US-10-267-255-76/c  
; Sequence 76, Application US/10267255  
; Publication No. US20030124689A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600 456US1  
; CURRENT APPLICATION NUMBER: US/10/267,255  
; PRIOR FILING DATE: 2002-10-09  
; PRIOR APPLICATION NUMBER: US 09/266,965  
; PRIOR FILING DATE: 1999-03-12  
; PRIOR APPLICATION NUMBER: US 08/624,447  
; PRIOR FILING DATE: 1996-08-19  
; PRIOR APPLICATION NUMBER: PCT/US94/11279  
; PRIOR FILING DATE: 1994-10-06  
; PRIOR APPLICATION NUMBER: US 08/133,963  
; PRIOR FILING DATE: 1993-10-07  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 76  
; LENGTH: 53500  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-10-267-255-76

Query Match 16.1%; Score 133.2; DB 15; Length 53500;  
Best Local Similarity 54.0%; Pred. No. 1,3e-27;  
Matches 295; Conservative 0; Mismatches 248; Indels 3; Gaps 1;

QY 15 CGCAACAATCGCAGAGGTGGGCAAGATGATGACCTGTACCGCGCTTGCTGAATC 74  
DB 7549 CTCACCGCTCCGAGAGATGGGCGGCTCTACGACCGGTTACCGCGCTGGAGCCG 7490  
QY 75 GGTGCGGGGCGGCGGCTGCGCATTCACACGCGCTACGAGAGACGAGGCGGCTTC 134  
DB 7489 CTCCTCGGAGAACCTGCACTTCGGCTACTGGAACT---CCCGACAGCAGGATGCC 7433  
QY 135 CTGGCAGCAGCGCGCGGCTCACCGACTTGTGCGGAAAGCAACGCTGCTCATG 194  
DB 7432 GCTGGCGAGGCCACCGACCGGCTCACGATGATGGCCGAGCGGCTGCGCATCGCGC 7373  
QY 195 CGGCGTTGCACTGATGATGGGGTGGGTTACCGGACACAGCGCTGCGCGTCCGCG 254  
DB 7372 CGGCTCCCGGCTCTGACCTCGGCTGCGGCGTGGGAGACCCCGGCGTACGATCGCCG 7313  
QY 255 CGAACAAGCGATCCAGATCACCGGATCACGTCAGCAGATGCAAGTGGCAATCGCGC 314  
DB 7312 GCTCAGCGAGCGCATGTACGGGCACTCGGTGAGGCAAGAGAGGTGCTCGGGGCA 7253  
QY 315 TGATGCGACGCGGAACGCGAATTAACCAACCGGTTGACTTCTGTCGTGATGCTCAT 374  
DB 7252 CGGCTGCGAGAGAGCGGCGGCTCGCGACCGGCGCGCTTCAGAGGCGGCGAGAT 7193  
QY 375 GTCCTGCGGATACCGGACCAATGTTTTCAGCGCGCGTGGGCAATGAGTGGCTTGA 434  
DB 7192 GAGCTTCCTTTGAGAGAGAGAGCTTCAGCGCGGATGAGCTTCGAAATGATCATCA 7133  
QY 435 GATGTCGAACCGGACCGATCCGGAATTCCTTGAATCTCAACCCCGTGGCAT 494  
DB 7132 CATGCCGAGCGGCGGAGGTGCTCGCGAGGTGCGCGGAGTGTGCGCGCGGCGCG 7073  
QY 495 CTTGGGCTACCGAGGTGCTCAAGAGAGAGCGGCGGCGGAGTGCCTGTCGAGGA 554  
DB 7072 TCTGTGCTCAACGACTTCTTCGAGCGGCGCGCGCTCGCCCGAGGAGGCGGCGCGT 7013  
QY 555 CAGGTG 560

DB 7012 CCAGCG 7007

RESULT 12  
US-09-980-217-1/c  
; Sequence 1, Application US/09980217  
; Publication No. US20040219645A1  
; GENERAL INFORMATION:  
; APPLICANT: Biotech Technology Limited  
; APPLICANT: Leadlay, Peter F  
; APPLICANT: Staunton, James  
; APPLICANT: Olynyk, Marko  
; TITLE OF INVENTION: Polyketides and their synthesis  
; FILE REFERENCE: IS/BP5858469  
; CURRENT APPLICATION NUMBER: US/09/980,217  
; PRIOR FILING DATE: 2002-05-06  
; PRIOR APPLICATION NUMBER: PCT/GB00/02072  
; PRIOR FILING DATE: 2000-05-30  
; PRIOR APPLICATION NUMBER: GB 9912563.5  
; PRIOR FILING DATE: 1999-05-28  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 30000  
; TYPE: DNA  
; ORGANISM: Streptomyces cinamomensis  
; FEATURE:  
; OTHER INFORMATION: Nucleotides 1 to 30000 of the monensin  
US-09-980-217-1

Query Match 14.5%; Score 120.2; DB 11; Length 30000;  
Best Local Similarity 52.5%; Pred. No. 5.8e-24;  
Matches 263; Conservative 0; Mismatches 238; Indels 0; Gaps 0;

QY 50 ACCTGTGACGCGGCTGTGAATCTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 109  
DB 10374 ACAAGCTTTCAGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 10315  
QY 110 ACTGGAGAAAGAGGCGGCGGCTTCTGCGACAGAGCGGCGGCGGCGGCGGCGGCT 169  
DB 10314 TCAGCGCGCGGCGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10255  
QY 170 TCGCGGAAGAGACCGGCTGATGAGGCGGCTTCACTGCTGATGAGGCGGCGGCGG 229  
DB 10254 TGATCGCGGCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 10195  
QY 230 GACAAACAGCGCTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 289  
DB 10194 GCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 10135  
QY 290 GCCAGGTGAAGTGGCATGCGGCTGATGAGGCGGCGGCGGCGGCGGCGGCGGCGG 349  
DB 10134 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 10075  
QY 350 TGACCTTCTGAGTGCATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 409  
DB 10074 TCGCTTTCAGAGTGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10015  
QY 410 CTTGGGCGGATGAGTGCATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 469  
DB 10014 GCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 9955  
QY 470 TTGAGTACTCAAAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 529  
DB 9954 ACCGCGTGCAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9895  
QY 530 GCGGCGGAGTCCGCGGCTCG 550  
DB 9894 ATCCAGCGGCGGCGGCGGCGG 9874

RESULT 13

```
US-10-437-963-72340/c
; Sequence 72340, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 72340
; LENGTH: 1773
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_72728C.1
US-10-437-963-72340
```

```
Query Match      11.3%; Score 93.8; DB 19; Length 1773;
Best Local Similarity 52.7%; Pred. No. 1.5e-16;
Matches 203; Conservative 0; Mismatches 182; Indels 0; Gaps 0;
```

```
OY      133  TCCTGGCAGACGCGCCGACCGGCTCACCGACCTTGTGCCGAACGACCGTGTCTGAT 192
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1268  TCCCAACCGAGGACGCGCTCCACGAGAGCGCGTCCGACCTCTCCAGGCCAAG 1209

OY      193  GCGCGCGTTTCAGCTGCTGATGTGGGATGCGGTACCGGACCAACGACGCTGGCGTCCG 252
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1208  CCGGACACCGGCTCTCGACGTGCGCTGCGCGTGGCGGCGGCGCATGCGCCATCGCC 1149

OY      253  CGCGCAACGCGCATCAGATCAACCGGATCAACCGTCAAGCAGGTGCAAGTGGCCATCGCC 312
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1148  GCCCACTCGGCTCCAGAGTGTGCGCATCAACATCAAGATACAGATGAAACCGGCC 1089

OY      313  GCTGATTTGGGACGCGGAACGCGGACTTAAGCAACGCGGTGAGACTTCTCGTGTGATGCC 372
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1088  CGCGGCAACCGGACGCGGCTCGACTCCCTCGAGGTGCTGCGGTAACTTC 1029

OY      373  ATGTCCCGCGGTACCGGACCAATGCTTTCAGCGCGGCTGGGCGCATGACGTGCTGTG 432
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1028  CTCTCAGCGGTTCTTCGAGCGCTCTTCGAGCGGCGCTACTCCATGAGGCCACCTGC 969

OY      433  GAGATGTCCGAACCGGACCGTGCATTCGCGGAAATCTTTCAGTACTCAAAACCGGTGAC 492
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      968  CACGCGCCCAAGGCTCAAGAGAGTCTTACGCGGAGGTCTTCCGGTGTCTCAAGCGGCGGC 909

OY      493  ATCTCGCGGTACCGAGGTGCTCA 517
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      908  CTCTACGTCTCTCAAGATGGGTCA 884
```

```
RESULT 14
US-10-259-194A-251
; Sequence 251, Application US/10259194A
; Publication No. US20040010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Ghassemlan, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
```

```
; APPLICANT: Provart, Nicholas
; APPLICANT: Rieke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
; CURRENT APPLICATION NUMBER: US/10/259,194A
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,743
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 662
; SOFTWARE: Patentlist.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 251
; LENGTH: 1092
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-259-194A-251
```

```
Query Match      11.3%; Score 93.2; DB 17; Length 1092;
Best Local Similarity 52.6%; Pred. No. 2.2e-16;
Matches 203; Conservative 0; Mismatches 183; Indels 0; Gaps 0;
```

```
OY      132  TTCCTGGCAGACGCGCCGACCGGCTCACCGACCTTGTGCCGAACGACCGTGTCTGA 191
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      321  TTCACCGCGAGCGCACGCGCGTCCACAGAGCGGCGTCCGACCTCTCCAGGCCAA 380

OY      192  TGGGCGCGTTTACTGCTGCTCAATGTGGGATGCGGTACCGGACCAACGACGCTGCCGTGCC 251
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      381  GCCGGACACCGCTCTCTCGACGTGCGGTGGCGGTGGCGGCGCATGCGCGCATCGC 440

OY      252  GCGGACACGCGATTCAGATCAACGGCATCACCGGATCACCGAGGTGCAAGTGGCATCGC 311
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      441  CGACACTCGCGCTCCAAAGTGTGCGCATACCATCAACAGATCAAGAGTGAACCGCGC 500

OY      312  CGTGTATTCGACGCGGAACGCGGACTTAAGCAACCGGATGACTTCTGTGCTGATGC 371
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      501  CCGGCGCAACCGCAAGCGGCGCTCGACTCCGCTGCGAGGTCTGTGCGGTAACTT 560

OY      372  CATGTCTCTGCGGTACCGGACATGCTTTCAGAGCGGCGCTGGCGCATGAGTGGCTGT 431
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      561  CCTTCAATGCGTCTTTCGAGCGCTCTTTCAGAGGCGGCTTAATCGAGGCGACCTG 620

OY      432  GGAATGTCCGAACCGGACCGTGCATTCGCGGAAATCTTTCAGTACTCAAAACCGGTG 491
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      621  CCAAGCGCCAGGCTGCAAGAGTCTTACGCGGAGGTCTTCCGCTGCTCAAGCGGCGC 680

OY      492  CATCTCGCGCTCAACGAGTGTCA 517
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      681  CCTTACGTCTCTACGAGTGGTCA 706
```

```
RESULT 15
US-10-259-194A-529
; Sequence 529, Application US/10259194A
; Publication No. US20040010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Ghassemlan, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Rieke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
```

;/ CURRENT APPLICATION NUMBER: US/10/259,194A  
;/ CURRENT FILING DATE: 2003-01-07  
;/ PRIOR APPLICATION NUMBER: US 60/325,277  
;/ PRIOR FILING DATE: 2001-09-26  
;/ PRIOR APPLICATION NUMBER: US 60/370,743  
;/ PRIOR FILING DATE: 2002-04-04  
;/ PRIOR APPLICATION NUMBER: US 60/370,620  
;/ PRIOR FILING DATE: 2002-04-04  
;/ NUMBER OF SEQ ID NOS: 662  
;/ SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta  
;/ SEQ ID NO 529  
;/ LENGTH: 1418  
;/ TYPE: DNA  
;/ ORGANISM: Trifolium aestivum  
US-10-259-194A-529

Query Match 10.8%; Score 89.4; DB 17; Length 1418;  
Best Local Similarity 52.2%; Pred. No. 2.6e-15;  
Matches 198; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 139 CAGCAGCCGCCGACCGGCTCACCGACCTTGTGCCCCGAAGACCGTGCTCGATGGCGGC 198  
DB 408 CCGGAGGCCACCGGGGTGACGAGAGCGGTGGCCGACCTGCTGAGGCGCGCCCGG 467  
QY 199 GTTCACTGCTCGATGTGGGGTGGGTCGAGACCAACGCGCTGCGCGCGCGAC 258  
DB 468 AAGCGCTGCTCGACGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 527  
QY 259 AAGCGATCCAGATCAACCGGATCACCGGTGAGCGAGGTGCAAGTGGCCATGCGCGCTGAT 318  
DB 528 TCCGGCTCCGAGCTGCGGATTAACATCAAGAGTCAACCGCGCGCGCGCGCGCG 587  
QY 319 TGGGACCGGAGACCGGACTAAGCCACCGGCTGACTTCTGCTGCTGATGCTATGCC 378  
DB 588 CACAAACCGGAGCGCGCTGAGCGGCAATGCGAGGTGTGTGCGCAACTTCATGGCC 647  
QY 379 CTGGCGTACCCGGACCAATGCTTTCAGCGCGCGCTGCGGCGCATGCGTGTGAGATG 438  
DB 648 ATGCGCTTCCGGAGCGCTTCTTCAGAGGCGCTTACTCCATCGAGGCGACCTGCCAGCG 707  
QY 439 TCCGAACCGGACCGGTGCATCCGGAAATCCTTCAGTACTCAAAACCGGTGGCATCTC 498  
DB 708 CCCAGGCTGAGAGACGCTTACGCGGAGGTGTTCGGGTGCTCAAGCGAGGCGGCTCTAC 767  
QY 499 GAGCGTACCGAGGTCTCA 517  
DB 768 GTCTCTACGAGTGGGTCA 786

Search completed: August 29, 2005, 17:23:43  
Job time : 718 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: August 29, 2005, 16:11:00 ; Search time 3885 Seconds  
(without alignments)  
10327.135 Million cell updates/sec

Title: US-10-069-353A-7

Sequence: 1 gtgtctgcaggtgtgcgcacc.....cgcggaagccggtcgcgtga 828

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4708233 seqs, 24227607955 residues

word size : 0

Total number of hits satisfying chosen parameters: 9416466

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

### Post-processing: Listing first 45 summaries

Database : GenEmbl : \*

```

1:  gb_ba:*
2:  gb_hng:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_sy:*
12: gb_ey:*
13: gb_un:*
14: gb_vi:*

```

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description
	1	828	100.0	828	6	AX089422	AX089422 Sequence	
C	2	828	100.0	45624	6	AX089419	AX089419 Sequence	
	3	828	100.0	50000	6	AX089416	AX089416 Sequence	
	4	828	100.0	80161	1	AY007564	AY007564 Saccharo	
	5	828	100.0	80161	6	AR165018	AR165018 Sequence	
	6	828	100.0	80161	6	BD137649	BD137649 Biosynth	
	7	828	100.0	80161	6	AR281866	AR281866 Sequence	
	8	57	6.9	36538	6	AX600587	AX600587 Sequence	
C	9	22	2.7	22	6	AX600621	AX600621 Sequence	
C	10	22	2.7	457	6	E07847	E07847 DNA sequenc	
C	11	22	2.7	457	6	E07848	E07848 DNA sequenc	
C	12	22	2.7	1410	6	E07846	E07846 DNA sequenc	
C	13	22	2.7	5998	1	AB032524	AB032524 Streptom	
C	14	22	2.7	12381	6	BD217469	BD217469 Polyketid	
C	15	22	2.7	12381	6	AX006889	AX006889 Sequence	
C	16	22	2.7	110000	1	AB017180_30	Continuation (31 c	
C	17	22	2.7	302675	1	AF005024	AF005024 Streptomy	
C	18	21	2.5	9251	1	AFR488769	AFR488769 Actinopla	
C	19	21	2.5	183847	5	AC145510	AC145510 Gasterost	

20	21	2.5	192556	5	AC145764	AC145764	Gasterost
21	20	2.4		6	AX600620	AX600620	Sequence
22	20	2.4	262	8	AY020544	AY020544	Oryza sat
C 23	20	2.4	833	8	R1CSPA	D67043	Oryza sativ
24	20	2.4	1578	8	D67043	AK101008	Oryza sat
25	20	2.4	1674	8	AK101008	PMAAT1	P.miliaceum
26	20	2.4	1677	8	PMAAT1	AK070560	Oryza sat
27	20	2.4	1765	8	AK070560	BD133827	Microbial
28	20	2.4	1832	6	BD133827	MTU40159	Mycobacteri
C 29	20	2.4	2700	6	MTU40159	AR053877	Sequence
C 30	20	2.4	3107	6	AR053877	AB070456	Rhodococc
31	20	2.4	5474	1	AB070456	AK092722	Sequence
32	20	2.4	8051	6	AR092722	Y11199	S.erythraea
33	20	2.4	8160	1	SEERGENE	AX000487	Sequence
34	20	2.4	8160	6	AX000487	BD073332	Gene for
35	20	2.4	8160	6	BD073332	AB054975	Terrabact
36	20	2.4	8179	1	AB054975	U77455	Saccharopol
37	20	2.4	8340	1	SEU77455	AB095015	Terrabact
38	20	2.4	11859	1	AB095015	AC005807	Mus muscu
39	20	2.4	65870	10	AC005807	CR522870_17	Continuatio
40	20	2.4	110000	1	AE000516	40	Continuatio
41	20	2.4	110000	1	AE000516	40	Continuatio
C 42	20	2.4	112907	10	AL670865	AL670865	Mouse DNA
C 43	20	2.4	135792	2	AP003544	AP003544	Oryza sat
C 44	20	2.4	148334	10	AY555278	AY555278	Mus muscu
45	20	2.4	149089	8	AP003991	AP003991	Oryza sat

## ALIGNMENTS

LOCUS	AX089422	828 bp	DNA	linear	PAT 21-MAR-2001
DEFINITION	Sequence 7 from Patent WO0116303.				
ACCESSION	AX089422				
VERSION	AX089422.1	GI:13443683			
KEYWORDS					
SOURCE					
ORGANISM	Saccharopolyspora spinosa Saccharopolyspora spinosa Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae, Pseudonocardiaceae, Saccharopolyspora.				
REFERENCE	1 Eberz, G., Moehrl, V., Froede, R., Velten, R. and Salas, J.A. Nucleic acids which code for the enzyme activities biosynthesis Patent: WO 0116303-A 7 08-MAR-2001; BAYER AG (DE)				
JOURNAL	Location/Qualifiers				
FEATURES	1..828 /organism="Saccharopolyspora spinosa" /mol_type="unassigned DNA" /db_xref="taxon:60894"				
Source	1..828 /note="unamed protein product; ORF1, O-Methyltransferase" /codon_start=1 /transf_table=11 /protein_id="CAC34874.1" /db_xref="GI:13443684" /translation="MLPGGAPTSQQVGQMYDLVTPLNSVAGSPCALHHGYWENDRR SMQCAADRLLDILVAERTVLDGVALTLCCTGCPALRVARDNAIQITGTVSQQVAV IAACADRGSHRVDSQVDAVMSLPYDNFADAMQSLLEMSPEPRALREILRVAF KPGGILVTEVVRKAGGMPVSGDRPTGRLCIAEQLLSLAAGEILDMEVDS RTRIFMQFAEELAHQHGLADRGPAVAGMAAVCDYEKVAHDMGYALITARKVVG				
CDS					
ORIGIN	Query Match 100.0%; Score 828; DB 6; Length 828; Best Local Similarity 100.0%; Pstd. No. 0; Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 GTGTTGCCAGGTGGCGGCACCAACATCGCAGAGGTTGGCGAGATGATGACCTGGTCACG 60 1 GTGTTGCCAGGTGGCGGCACCAACATCGCAGAGGTTGGCGAGATGATGACCTGGTCACG 60				
DB	1 GTGTTGCCAGGTGGCGGCACCAACATCGCAGAGGTTGGCGAGATGATGACCTGGTCACG 60				

QY 61 CCGTTCGTAACCTCGGTCCGCGGCGCCCTGCGCCATCCACACCGGCTACTGGAGAAC 120  
DB 61 CCGTTCGTAACCTCGGTCCGCGGCGCCCTGCGCCATCCACACCGGCTACTGGAGAAC 120  
QY 121 GACGGCGCGGCTTCTCTGGACGACAGGCGCGGACCGGCTCAACCGACTTGTCCGGAACG 180  
DB 121 GACGGCGCGGCTTCTCTGGACGACAGGCGCGGACCGGCTCAACCGACTTGTCCGGAACG 180  
QY 181 ACCGCTCGATGAGCGGCGGCTGGAATGCTGATGAGGAGGAGGAGGAGGAGGAGGAGG 240  
DB 181 ACCGCTCGATGAGCGGCGGCTGGAATGCTGATGAGGAGGAGGAGGAGGAGGAGGAGG 240  
QY 241 CTGCGCGTCCGCGCGACACACCGGATCCAGATCAACCGGATCAACCGGATCAACCGGATCA 300  
DB 241 CTGCGCGTCCGCGCGACACACCGGATCCAGATCAACCGGATCAACCGGATCAACCGGATCA 300  
QY 301 GTGGCCATGCGCGCTGATGTCGCAACCGGACCGGACCGGACCGGACCGGACCGGACCGG 360  
DB 301 GTGGCCATGCGCGCTGATGTCGCAACCGGACCGGACCGGACCGGACCGGACCGGACCGG 360  
QY 361 TGGGTCGATGCGATGCTGCGCTGCGGACCGGACCGGACCGGACCGGACCGGACCGGACG 420  
DB 361 TGGGTCGATGCGATGCTGCGCTGCGGACCGGACCGGACCGGACCGGACCGGACCGGACG 420  
QY 421 CAGTCGCTGTTGAGATGTCGCAACCGGACCGGACCGGACCGGACCGGACCGGACCGGACG 480  
DB 421 CAGTCGCTGTTGAGATGTCGCAACCGGACCGGACCGGACCGGACCGGACCGGACCGGACG 480  
QY 481 AAACCCGCGTGCATCTCGGCGCTCAACCGGATGCTCAACCGGATGCTCAACCGGATGCTCA 540  
DB 481 AAACCCGCGTGCATCTCGGCGCTCAACCGGATGCTCAACCGGATGCTCAACCGGATGCTCA 540  
QY 541 CCGGTCGCGGAGACAGTGGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACG 600  
DB 541 CCGGTCGCGGAGACAGTGGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACG 600  
QY 601 GAATGCTGCTGACGCGGAGGTTGAGATCTCGATGAGGAGAGAGGAGGAGGAGGAGGAGG 660  
DB 601 GAATGCTGCTGACGCGGAGGTTGAGATCTCGATGAGGAGAGAGGAGGAGGAGGAGGAGG 660  
QY 661 CCGTACTTATGATCCGCAATTCGCGGAGAGGCTGCTGCGCACACGAGGATGCGGAGC 720  
DB 661 CCGTACTTATGATCCGCAATTCGCGGAGAGGCTGCTGCGCACACGAGGATGCGGAGC 720  
QY 721 AGGTACGGGCGGAGTGTGCGCGGCTGCGGCGGCGGCGGCTGCGGATTAAGAATAATGCG 780  
DB 721 AGGTACGGGCGGAGTGTGCGCGGCTGCGGCGGCGGCGGCTGCGGATTAAGAATAATGCG 780  
QY 781 CACGACATGGGCTATGCGATTCTGACGCGGCGGAAAGCCGGTGGCTGA 828  
DB 781 CACGACATGGGCTATGCGATTCTGACGCGGCGGAAAGCCGGTGGCTGA 828

RESULT 2  
AX089419/c 45624 bp DNA linear PAT 21-MAR-2001  
LOCUS Sequence 4 from Patent WO0116303.  
DEFINITION AX089419  
ACCESSION AX089419  
VERSION AX089419.1 GI:13443680  
KEYWORDS  
SOURCE  
ORGANISM  
Saccharopolyspora spinosa  
Saccharopolyspora spinosa  
Bacteria; Actinobacteria; Actinobacteriales;  
Pseudonocardiales; Pseudonocardiales; Saccharopolyspora.  
REFERENCE  
1 Ebertz,G., Moehrl,V., Froede,R., Velten,R. and Salas,J.A.  
Nucleic acids which code for the enzyme activities of the spinosyn  
biosynthesis  
JOURNAL Patent: WO 0116303-A 4 08-MAR-2001;  
BAYER AG (DE)  
FEATURES  
location/Qualifiers  
1..45624  
source

ORIGIN  
Query Match 100.0%; Score 828; DB 6; Length 45624;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
/organism="Saccharopolyspora spinosa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:60894"

QY 1 GTGTTGCCAGGTGGGACCAACATCCGACGAGGTTGGGAGATGATGACTTGGTCAACG 60  
DB 828 GTGTTGCCAGGTGGGACCAACATCCGACGAGGTTGGGAGATGATGACTTGGTCAACG 769  
QY 61 CCGTTCGTAACCTCGGTCCGCGGCGCCCTGCGCCATCCACACCGGCTACTGGAGAAC 120  
DB 768 CCGTTCGTAACCTCGGTCCGCGGCGCCCTGCGCCATCCACACCGGCTACTGGAGAAC 709  
QY 121 GACGGCGCGGCTTCTCTGGACGACAGGCGCGGACCGGCTCAACCGACTTGTCCGGAACG 180  
DB 708 GACGGCGCGGCTTCTCTGGACGACAGGCGCGGACCGGCTCAACCGACTTGTCCGGAACG 649  
QY 181 ACCGCTCGATGAGCGGCGGCTGGAATGCTGATGAGGAGGAGGAGGAGGAGGAGGAGG 240  
DB 648 ACCGCTCGATGAGCGGCGGCTGGAATGCTGATGAGGAGGAGGAGGAGGAGGAGGAGG 589  
QY 241 CTGCGCGTCCGCGCGACACACCGGATCCAGATCAACCGGATCAACCGGATCAACCGGATCA 300  
DB 588 CTGCGCGTCCGCGCGACACACCGGATCCAGATCAACCGGATCAACCGGATCAACCGGATCA 529  
QY 301 GTGGCCATGCGCGCTGATGTCGCAACCGGACCGGACCGGACCGGACCGGACCGGACCGG 360  
DB 528 GTGGCCATGCGCGCTGATGTCGCAACCGGACCGGACCGGACCGGACCGGACCGGACCGG 469  
QY 361 TGGGTCGATGCGATGCTGCGCTGCGGACCGGACCGGACCGGACCGGACCGGACCGGACG 420  
DB 468 TGGGTCGATGCGATGCTGCGCTGCGGACCGGACCGGACCGGACCGGACCGGACCGGACG 409  
QY 421 CAGTCGCTGTTGAGATGTCGCAACCGGACCGGACCGGACCGGACCGGACCGGACCGGACG 480  
DB 408 CAGTCGCTGTTGAGATGTCGCAACCGGACCGGACCGGACCGGACCGGACCGGACCGGACG 349  
QY 481 AAACCCGCGTGCATCTCGGCGCTCAACCGGATGCTCAACCGGATGCTCAACCGGATGCTCA 540  
DB 348 AAACCCGCGTGCATCTCGGCGCTCAACCGGATGCTCAACCGGATGCTCAACCGGATGCTCA 289  
QY 541 CCGGTCGCGGAGACAGTGGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACG 600  
DB 288 CCGGTCGCGGAGACAGTGGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACG 229  
QY 601 GAATGCTGCTGACGCGGAGGTTGAGATCTCGATGAGGAGAGGAGGAGGAGGAGGAGGAGG 660  
DB 228 GAATGCTGCTGACGCGGAGGTTGAGATCTCGATGAGGAGAGGAGGAGGAGGAGGAGGAGG 169  
QY 661 CCGTACTTATGATCCGCAATTCGCGGAGAGGCTGCTGCGCACACGAGGATGCGGAGC 720  
DB 168 CCGTACTTATGATCCGCAATTCGCGGAGAGGCTGCTGCGCACACGAGGATGCGGAGC 109  
QY 721 AGGTACGGGCGGAGTGTGCGCGGCTGCGGCGGCGGCTGCGGATTAAGAATAATGCG 780  
DB 108 AGGTACGGGCGGAGTGTGCGCGGCTGCGGCGGCGGCTGCGGATTAAGAATAATGCG 49  
QY 781 CACGACATGGGCTATGCGATTCTGACGCGGCGGAAAGCCGGTGGCTGA 828  
DB 48 CACGACATGGGCTATGCGATTCTGACGCGGCGGAAAGCCGGTGGCTGA 1

RESULT 3  
AX089416 50000 bp DNA linear PAT 21-MAR-2001  
LOCUS AX089416  
DEFINITION Sequence 1 from Patent WO0116303.  
ACCESSION AX089416  
VERSION AX089416.1 GI:13443677  
KEYWORDS

SOURCE Saccharopolyspora spinosa  
 ORGANISM Saccharopolyspora spinosa  
 Bacteria; Actinobacteridae; Actinomycetales;  
 Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.

REFERENCE 1  
 Eberz G., Moehle V., Froede R., Velten R. and Salas J.A.  
 Nucleic acids which code for the enzyme activities of the spinosyn  
 biosynthesis  
 Patent: WO 0116303-A 1 08-MAR-2001;

JOURNAL  
 BAYER AG (DE)  
 Location/Qualifiers  
 1..50000  
 /organism="Saccharopolyspora spinosa"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:60894"

ORIGIN

Query Match 100.0%; Score 828; DB 6; Length 50000;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTTCCAGTGGCCGACCAACATGCGAGGTTGGGAGATGATGACTTGTCAGC 60  
 DB 44797 GTGTTCCAGTGGCCGACCAACATGCGAGGTTGGGAGATGATGACTTGTCAGC 44856  
 QY 61 CCGTTGCTGAATCTCGGTCCGGGCGGCGCCCTGCGCATCCACACGCGCTACTGGAGAAC 120  
 DB 44857 CCGTTGCTGAATCTCGGTCCGGGCGGCGCCCTGCGCATCCACACGCGCTACTGGAGAAC 44916  
 QY 121 GAGGGCGGGCTTCCTGGAGAGAGGCGCGGCGGCTGACCGACCTTGTCGCGGAGACG 180  
 DB 44917 GAGGGCGGGCTTCCTGGAGAGAGGCGCGGCGGCTGACCGACCTTGTCGCGGAGACG 44976  
 QY 181 ACCGTCTCGATGGCGGCGGTTCGACTGCTGATGTGGGGTGGCGATACCGGACCAACGCG 240  
 DB 44977 ACCGTCTCGATGGCGGCGGTTCGACTGCTGATGTGGGGTGGCGATACCGGACCAACGCG 45036  
 QY 241 CTGCGCGTCCGGCGGCGCAACCGGATCCAGATCAACCGGCTACCGTCAAGCGGTGCA 300  
 DB 45037 CTGCGCGTCCGGCGGCGCAACCGGATCCAGATCAACCGGCTACCGTCAAGCGGTGCA 45096  
 QY 301 GTGGCCATGCGCGCTGATGGGCAACCGGAGCGGGAATGACCGGCGGTGCACTTCG 360  
 DB 45097 GTGGCCATGCGCGCTGATGGGCAACCGGAGCGGGAATGACCGGCGGTGCACTTCG 45156  
 QY 361 TGGGTGATGATGCTGCTGCGGTACCGGCAATGCTTGCAGCGCGCTGCGCATG 420  
 DB 45157 TGGGTGATGATGCTGCTGCGGTACCGGCAATGCTTGCAGCGCGCTGCGCATG 45216  
 QY 421 CAGTGGCTGTGGAGATGTCGAAACCGGACCTGTGCCATCCGGGAATCTTTCAGATGTC 480  
 DB 45217 CAGTGGCTGTGGAGATGTCGAAACCGGACCTGTGCCATCCGGGAATCTTTCAGATGTC 45276  
 QY 481 AAACCCGGTGGCATCTCGGCGGTGACCGGAGTGTCAAGAGAGCGGCGGCGGAGT 540  
 DB 45277 AAACCCGGTGGCATCTCGGCGGTGACCGGAGTGTCAAGAGAGCGGCGGCGGAGT 45336  
 QY 541 CCGGTGTCGGGGGACAGGTGGCGGACCGGCTTTCGATGCTGCTGCTGCAACTTCTG 600  
 DB 45337 CCGGTGTCGGGGGACAGGTGGCGGACCGGCTTTCGATGCTGCTGCTGCAACTTCTG 45396  
 QY 601 GAATCGCTGTGGAGCGGGGTTGAGATCTCGATTGGGAGAGCGTGTGTGAGGAC 660  
 DB 45397 GAATCGCTGTGGAGCGGGGTTGAGATCTCGATTGGGAGAGCGTGTGTGAGGAC 45456  
 QY 661 CGGACTTCATGCGCGGAGTTGCGCGAGAGCTGCTGCTGCGGACCGGAGTGGGAG 720  
 DB 45457 CGGACTTCATGCGCGGAGTTGCGCGAGAGCTGCTGCTGCGGACCGGAGTGGGAG 45516  
 QY 721 AGGATCGGCGCGGCTGTGCGCGCTGGGCGCGCGGCTTTCGATGCTGCAATATATGCC 780  
 DB 45517 AGGATCGGCGCGGCTGTGCGCGCTGGGCGCGCGGCTTTCGATGCTGCAATATATGCC 45576

QY 781 CACGACATGGGCTATGCGATTCTGACGGCGCGGAGACCGGCTGCGCTGA 828  
 DB 45577 CACGACATGGGCTATGCGATTCTGACGGCGCGGAGACCGGCTGCGCTGA 45624

RESULT 4  
 AY007564  
 LOCUS  
 DEFINITION Saccharopolyspora spinosa probable lysR-type transcriptional  
 regulator, probable keto acyl reductase, spinosad biosynthetic gene  
 cluster, complete sequence, and probable exodeoxyribonuclease V  
 genes, complete cds, and unknown gene.

ACCESSION  
 VERSION AY007564.1 GI:13162633  
 KEYWORDS  
 SOURCE Saccharopolyspora spinosa  
 ORGANISM Saccharopolyspora spinosa  
 Bacteria; Actinobacteridae; Actinomycetales;  
 Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.

REFERENCE 1 (bases 1 to 80161)  
 Waldron, C., Matsushima, P., Rosteck, P. R. Jr., Broughton, M. C.,  
 Turner, J., Madduri, K., Crawford, K. P., Merlo, D. J. and Baltz, R. H.  
 Cloning and analysis of the spinosad biosynthetic gene cluster of  
 Saccharopolyspora spinosa  
 Chem. Biol. 8 (5), 487-499 (2001)

JOURNAL  
 MEDLINE 21257765  
 PUBMED 11358695  
 REFERENCES 1 (bases 1 to 80161)  
 Waldron, C., Matsushima, P., Rosteck, P. R. Jr., Broughton, M. C.,  
 Turner, J., Madduri, K., Crawford, K. P., Merlo, D. J. and Baltz, R. H.  
 Direct Submission  
 Submitted (18-AUG-2000) Dow Agrosciences, 9330 Zionsville Rd.,  
 Indianapolis, IN 46268, USA

TITLE  
 JOURNAL  
 Indianapolis, IN 46268, USA

FEATURES  
 source  
 1..80161  
 Location/Qualifiers  
 /organism="Saccharopolyspora spinosa"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:60894"  
 complement(1135..1971)  
 /note="ORF L16"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable lysR-type transcriptional regulator"  
 /protein\_id="AAC23282.1"  
 /db\_xref="GI:13162633"  
 /translation="METRELRYFAVAEELHFGRAAQRIGIAQPPISRTIAQEQRLG  
 VILQRTSRKVSILTEAGAMLTEGRAIILGALAAARRTQRASTQPSVLAKAGASG  
 ELLAKLIDAYAPGAVADILCESQPKTHIDGADVALIHOPEDPYAEIDIELN  
 TEOOVAIILPTSHPLASEPHVMADVSLPDLPLAWPGDGVYPDGPGVEVNRQTLF  
 QMIALGRITVWNPBESSRVNLEGLAAVPLDAPDVTTVIAMPPHSRSLAAGLVAVAT  
 LI"  
 2024..2791  
 /note="ORF L15"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable keto acyl reductase"  
 /protein\_id="AAC23281.1"  
 /db\_xref="GI:13162653"  
 /translation="MVLVRRFRFRATLESNEQTALVTGANKIGIYEIAGIGALGWS  
 VGIAGARDHQRGDDAVAKLRADQVDAFAVSLDVTDDASVAAAALLEERAGRDVLVN  
 AGIAGAMPBEPSTVPASLRVAVETNIGVETVTAMLPILRRSERPRIWVNSHAS  
 LTIOTTPGVLDIGISGAVSRKTFINAIITIOYAKELSDPINIKNNACPEYVATDLNGF  
 HGTSTPADGARAIARLATIPDDGPTGMDDDAGNVW"  
 complement(3416..4165)  
 /gene="spnS"  
 complement(3416..4165)  
 /note="probably involved in fosfomycin biosynthesis"  
 /gene="spnS"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="probable N-dimethyltransferase"  
 /protein\_id="AAC23280.1"

gene  
CDS  
/db\_xref="GI:13162652"  
/translation="MSRVSDFPASTSVSPDHADIDVAIHSARGRDWAAGEVVOI  
VTRLPFAOSLIDVACGTAHLEPRRAEYAKAGLESDAMEEIIARRVPEVPIHIG  
IRDPDLGEPDVIYTCCTPTAAMRTVVDLRRYTRMMARHLAPGSAVIEPWPPEKFI  
DGFVGAVAHGESEVLSRSHSVLEGRTSRMTVRYTVAPETGRDTEFEILSTED  
EYTALEAGIAEAEVLPAPNGRGLFVGIRN"  
complement (4168. .5325)  
/gene="spnr"  
complement (4168. .5325)  
/gene="spnr"  
complement (4168. .5325)  
/note="probably involved in forosamine biosynthesis"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="probable aminotransferase"  
/protein\_id="AAG23279.1"  
/db\_xref="GI:13162651"  
/translation="MINLHQPILGTBEBDAIEAVPASNWIGLPRTTPEAEFAHILG  
VDPEQVPLNSTALPLTVQVLDLGPDDVVLPEISFVAANALIASGARPFCDVD  
PRTNPLDDVARIATTPATKAVILHYGSSPGEVTAIDFCSEKGLMIEDSACAVAS  
SVHGACGTFGDLATWSPDMKILVTGOGMFAADPELAHARARLAYHLEOMSGFD  
SAKSNRMWDIREDICORLIGDMPTALGSYQARKPEINRPREIATQYDRLISDY  
PEVLLPPTLPDGHVSSHFYVQVLAPEIRDOVQOMLRGCIYTSRYRPLKVPITYRA  
DCKLPASADACRRTLLPLHPSLDDAEVRYADEFOKAVERHHISQSSPLRK"  
complement (5363. .6751)  
/gene="spnq"  
complement (5363. .6751)  
/gene="spnq"  
complement (5363. .6751)  
/note="probably involved in forosamine biosynthesis"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="probable NDP-hexose-3,4-dehydratase"  
/protein\_id="AAG23278.1"  
/db\_xref="GI:13162650"  
/translation="MQSRKTRALGKGARVTSQDDTCTATBMPDADKORILASVRYD  
HREQSPFVAGSTPIRPSGAVLDEDDVALEALELRIAAGNARFEESEFAVFG  
LRKALVNSGSSANLALSILTPSLGEARLPGEDEVITAAVGFPTTINPAVQNLV  
VFVDELGTYNATPDIRIKAATERTRAIMLANTLGNPPADEIAIEAKHEHFLVEDN  
CDVAGSTRGLTGRFDLTVSEFYBAHHITSGEGCVLTGSLAARIIEISLRMGSD  
CKCEGVNTQCKRPDYHIGTLPCGYDHKTPSHGCVNLKTTDQALALQSLTISA  
FGSARRRWRRLREBLSGLPGILLVATPHSDPSFGFATISADAGFTRALVNFLE  
VSNIGTRLPFGNITRHPAFEQVRYIADALTNDSIYDRTFWGVYGTIDQMLDYV  
VESIAEFAKSS"  
7083. .8450  
/gene="spnr"  
7083. .8450  
/gene="spnr"  
/note="involved in forosamine addition"  
/codon\_start=1  
/transl\_table=11  
/product="probable NDP-forsamy1transferase"  
/protein\_id="AAG23277.1"  
/db\_xref="GI:13162649"  
/translation="MILGMLPGCSIAIGEFMRVLEFTPLPASHFENVLVLAMALPAAG  
HEVRVAICPNMYSWTGAGLTAVPGDELIDISLAKNELVIGSGVSEDEGRHPELF  
DELSINGRDIDAEQLAVDDRDLGFAEKMOPLDVMDAMVCSGQVAVRVALG  
ARHVMLVALDYSGLRSGFLPEYDSEKSPBOVDPLGWLAKLAFGATDEBIVTG  
QATIDPISMRALPYDLIDISMRFYPYNGPVLPEMIAERPKPVCTRGILTKRRLS  
RTBQYGEOSDEQAMERLIRGAARLVEVYATISDDVEBRMBELPSNVVHEVTLR  
NELLESCVLIHHGSTTQETATVNGVQPLLPGTFWESRRAEILLARAGLVLDPA  
TFTEDVDVQOLARLDEPSFANANALIRREIESSPHDIDVRLKLVAREENRTQG  
SDGHP"  
8967. .10427  
/gene="spno"  
8967. .10427  
/gene="spno"  
/note="involved in forosamine biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="probable NDP-hexose-2,3-dehydratase"  
/protein\_id="AAG23276.1"  
/db\_xref="GI:13162648"

gene  
CDS  
/translation="MSSSVEAEASAAAPLGSNTRRPVDSALSACNGMIPTTEPHCWL  
ADRGENSPENRRLPDRLSKWKCDASTENMLVHADGPPTVEGLOVENTYGAASMQ  
PINOACGFLIGLIVKRLIDVTLHCLMGEKPGVNNVLTCLSPYQARNTVGAOAHRS  
VPRVYDFLHGRGRVLDVLIQSCQSGFWFKRRKRNMYEQEEVPLVPCMTLLQO  
VLALRDONTVMKOTRTVLVSTCPHDSATGEBLASBEPKQAVARLSHSDISSIS  
EAVWFEBEAKARYLRATRVPLSVDSKMYRDTBIAHQDKYEFVAVIVASATREVA  
SWTQPMIEPREQGEIALLVKRIGVILHGLVHARAEAGKMTAEIAPVQGSVANYOST  
PSNDMPFLDVLTADEIVAYESIIEEGGRFYQAOQRVRIIEVHDFARPPSDFR  
MWTIGQGEILRSTHPLNIQARSLVVALSHSLMALGR"  
10436. .11434  
/gene="spnr"  
10436. .11434  
/gene="spnr"  
/note="involved in forosamine biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="probable NDP-hexose-3-ketoreductase"  
/protein\_id="AAG23275.1"  
/db\_xref="GI:13162647"  
/translation="MRKPRRIGVLGCSAFWRMLPAMCDVAETEVVAVASRDPKAE  
RFAARPCBAIVGYQRLERPDIIDAVVPLPFGMHAEMIGALBADKHVLAKEPLTTT  
ASDTRIVGLARKNLILRENYLFIHGRHDVVDLLOSGEIGLEFETVFGIPLP  
DIDIRITELGGGLDITGVPAHANHFLGLVTVGASHEAQESGVDSGVLLQ  
SEGGTVAHLGFEVHYRSAVELWSSRGRLVDBAFPPAPWQAVIRIERGVVDELS  
LPAEDQVKATVAFARDIRAGTGVDDPAVAGDSGSMIQOALVEALGOARRCGST"  
complement (11530. .12492)  
/gene="spnm"  
complement (11530. .12492)  
/gene="spnm"  
complement (11530. .12492)  
/note="involved in spinosyn aglycone biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="Spnm"  
/protein\_id="AAG23274.1"  
/db\_xref="GI:13162646"  
/translation="WENAVSGTVLVENIPMPREDRPIITRAVGTGHLGSOVABSYLIR  
TGTEPERTELAVALDRGMAVITDYBELGPTHTTVVGAQGANMLDAARAQRLPG  
SGLTDDCPVIGWGAQGOASAFGEHLPYVAPELRIRAAAGAVPIDLDIDITHRNG  
VFTPGLVAVGWAAAYPDLFPDELTEAGRTAVDVBELGAPELVTFGRLESDPL  
DTSLFPPORRRAIRLAVSVAINGRPVPLTVNHSSTDDELVPVFGERLDSYAAOT  
PVNRHPLSGLAHPAALASSRVVYVSWDHFSEBSAISGRDAR"  
complement (12696. .13547)  
/gene="spnl"  
complement (12696. .13547)  
/gene="spnl"  
/note="involved in spinosyn aglycone biosynthesis;  
possibly involved in polyketide bridging"  
/codon\_start=1  
/transl\_table=11  
/product="methyltransferase-1like protein"  
/protein\_id="AAG23273.1"  
/db\_xref="GI:13162645"  
/translation="MESIPDALHAGRLPHHGVYAGRYEDAGATPWSADADOLDLFI  
DKALRPGAHLPDLGCGNGOPVVRACASGVRVTGTVNKHOLAATRLANETGLAS  
LEFDVDAQLPYDGFQOAMAMQSVQIVDOAAALREVRILPEGRVLDGITR  
VRLPEEYAAVWTGTATHTLSPTLVEAEFELIEVLDLPAQTRCMWSVWVDELRLT  
DELAGFPAVAVGTQGYLDIAAKHGPAQIILAAVAYEKHPDYARNESMGFMLL  
QARKKQ"  
complement (13592. .14785)  
/gene="spnk"  
complement (13592. .14785)  
Query Match 100.0%; Score 828; DB 1; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GTGTTGCCAGGTGGCGCAACAATCGACAGGTTGGCGAGATGTATGACTGTGCAAG 60  
|||  
Db 20168 GTGTTGCCAGGTGGCGCAACAATCGACAGGTTGGCGAGATGTATGACTGTGCAAG 20227  
|||  
Qy 61 CCGTTGCTGAACCTGGTCGGGGGGGGCCCTGCGCCCATCCACCAAGGCTACTGGGAAC 120  
|||  
Db 20228 CCGTTGCTGAACCTGGTCGGGGGGGGCCCTGCGCCCATCCACCAAGGCTACTGGGAAC 20287  
|||

ORIGIN		
Query Match	100.0%	Score 828; DB 6; Length 80161;
Best Local Similarity	100.0%	Pred. No. 0;

REFERENCE  
AUTHORS  
1 (bases 1 to 80161)  
Baltz, R.H., Broughton, M.C., Crawford, K.P., Madduri, K., Merlo, D.J.,  
Treadway, P.J., Turner, J.R. and Waldron, C.

TITLE Biosynthetic genes for spinosyn insecticide production  
JOURNAL Patent: JP 2002505881-A 1 26-FEB-2002;  
COMMENT OS Saccharopolyspora spinosa  
PN JP 2002505881-A/1  
PD 26-FEB-2002  
PF 16-FEB-1998 JP 2000535754  
PR 09-MAR-1998 US 09/036987  
PI RICHARD H BALTZ, M CHRISTINE BROUGHTON, KATHRYN P CRAWFORD, PI  
KRISHNAMURTHY MADHURI, DONALD J MERLO, PATTI J TREADWAY, JUAN R PI  
TURNER  
PI CLIVE WALDRON  
PC C12N1/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12P19/62, C12Q1/  
PC 68//  
PC C07K14/195, C12N15/00, C12N5/00  
CC Biosynthetic genes for spinosyn insecticide production FH  
CY Key Location/Qualifiers  
FT source 1. .80161  
FEATURES Location/Qualifiers  
source 1. .80161  
/organism="Saccharopolyspora spinosa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:60894"  
ORIGIN  
Query Match 100.0%; Score 828; DB 6; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 1 GTGTTGCCAGGTGGCGCACCAATCGACAGAGTTGGGACAGATGTATGACTGTGTACG 60  
Db 20168 GTGTTGCCAGGTGGCGCACCAATCGACAGAGTTGGGACAGATGTATGACTGTGTACG 20227  
CY 61 CCGTTGCTGAATCGGTGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120  
Db 20228 CCGTTGCTGAATCGGTGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 20287  
CY 121 GACGGCGGGGCTTCCTGGCAGCAGGCGCGCCGACCGGCTCACCGACCTTGTGCCCGAACG 180  
Db 20288 GACGGCGGGGCTTCCTGGCAGCAGGCGCGCCGACCGGCTCACCGACCTTGTGCCCGAACG 20347  
CY 181 ACCGTGCTGATGGCGGGGGTTCGACTGCTCGATGTGGGGTGGCGTAACCGACACCAACG 240  
Db 20348 ACCGTGCTGATGGCGGGGGTTCGACTGCTCGATGTGGGGTGGCGTAACCGACACCAACG 20407  
CY 241 CTGGCGGTGGCGCGGACGACGCGATCGAGATCAACCGGATCAACGTCAGCAGGTTGCA 300  
Db 20408 CTGGCGGTGGCGCGGACGACGCGATCGAGATCAACCGGATCAACGTCAGCAGGTTGCA 20467  
CY 301 GTGGCCATCGCGCTGATTTGGCGCAGCGAACCGGACTTAAGCCACCGGGTGGACTTCTCG 360  
Db 20468 GTGGCCATCGCGCTGATTTGGCGCAGCGAACCGGACTTAAGCCACCGGGTGGACTTCTCG 20527  
CY 361 TGGCTGATGCTATGCTCTCTGCGGTAACCGGACATATGCTTTTGAAGCGCGCTGGGCCATG 420  
Db 20528 TGGCTGATGCTATGCTCTCTGCGGTAACCGGACATATGCTTTTGAAGCGCGCTGGGCCATG 20587  
CY 421 CAGTGCCTGTTGAGATGTCGAAACCGGACCGGTCATCCGGGAAATCTTTCGAGTACTC 480  
Db 20588 CAGTGCCTGTTGAGATGTCGAAACCGGACCGGTCATCCGGGAAATCTTTCGAGTACTC 20647  
CY 481 AAACCCGGTGGATCTCTCGCGTCAACGAGGTGTCMAACGAAACGAGCGCGCGCGAGT 540  
Db 20648 AAACCCGGTGGATCTCTCGCGGTCAACGAGGTGTCMAACGAAACGAGCGCGCGAGT 20707  
CY 541 CCGGTGTCCGGGACAGGTGGCGACCGGCTTCCGATCTGCTGGCTGAGCACTTCTG 600  
Db 20708 CCGGTGTCCGGGACAGGTGGCGACCGGCTTCCGATCTGCTGGCTGAGCACTTCTG 20767  
CY 601 GAATGCTGCTGACGCGGGGTTGAGATTCCTCGATTGGGAGAGATGTTCGCGAGGAC 660  
Db 20768 GAATGCTGCTGACGCGGGGTTGAGATTCCTCGATTGGGAGAGATGTTCGCGAGGAC 20827

661 CCGTACTTCATGCGCAGATTGCGCGAAGAGCTCGTCGACACGACGCGATCGCGAC 720  
20828 CCGTACTTCATGCGCAGATTGCGCGAAGAGCTCGTCGACACGACGCGATCGCGAC 20887  
CY 721 AGGTACGGGCGCGCTGTGCGCGGCTGGGCGCGCGCTGCTCCGATTTATGAAATATGCC 780  
Db 20888 AGGTACGGGCGCGCTGTGCGCGGCTGGGCGCGCGCTGCTCCGATTTATGAAATATGCC 20947  
CY 781 CACGACATGGGCTATGAGATTCTGACGCGCGGGAAGCCGGTCGAGTGA 828  
Db 20948 CACGACATGGGCTATGAGATTCTGACGCGCGGGAAGCCGGTCGAGTGA 20995  
RESULT 7  
LOCUS AR281866  
DEFINITION Sequence 1 from patent US 6521406.  
ACCESSION AR281866  
VERSION AR281866.1 GI:29717767  
KEYWORDS  
SOURCE .  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 80161)  
AUTHORS Baltz, R.H., Broughton, M.C., Crawford, K.P., Madhuri, K., Merlo, D.J.,  
Treadway, P.J., Turner, J.R. and Waldron, C.  
TITLE Spig, a gene for spinosyn insecticide biosynthesis  
JOURNAL Patent: US 6521406-A 1 18-FEB-2003;  
FEATURES Location/Qualifiers  
source 1. .80161  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 100.0%; Score 828; DB 6; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 1 GTGTTGCCAGGTGGCGCACCAATCGACAGAGTTGGGACAGATGTATGACTGTGTACG 60  
Db 20168 GTGTTGCCAGGTGGCGCACCAATCGACAGAGTTGGGACAGATGTATGACTGTGTACG 20227  
CY 61 CCGTTGCTGAATCGGTGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120  
Db 20228 CCGTTGCTGAATCGGTGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 20287  
CY 121 GACGGCGGGGCTTCCTGGCAGCAGGCGCGCCGACCGGCTCACCGACCTTGTGCCCGAACG 180  
Db 20288 GACGGCGGGGCTTCCTGGCAGCAGGCGCGCCGACCGGCTCACCGACCTTGTGCCCGAACG 20347  
CY 181 ACCGTGCTGATGGCGGGGGTTCGACTGCTCGATGTGGGGTGGCGTAACCGACACCAACG 240  
Db 20348 ACCGTGCTGATGGCGGGGGTTCGACTGCTCGATGTGGGGTGGCGTAACCGACACCAACG 20407  
CY 241 CTGGCGGTGGCGCGGACGACGCGATCGAGATCAACCGGATCAACGTCAGCAGGTTGCA 300  
Db 20408 CTGGCGGTGGCGCGGACGACGCGATCGAGATCAACCGGATCAACGTCAGCAGGTTGCA 20467  
CY 301 GTGGCCATCGCGCTGATTTGGCGCAGCGAACCGGACTTAAGCCACCGGGTGGACTTCTCG 360  
Db 20468 GTGGCCATCGCGCTGATTTGGCGCAGCGAACCGGACTTAAGCCACCGGGTGGACTTCTCG 20527  
CY 361 TGGCTGATGCTATGCTCTCTGCGGTAACCGGACATATGCTTTTGAAGCGCGCTGGGCCATG 420  
Db 20528 TGGCTGATGCTATGCTCTCTGCGGTAACCGGACATATGCTTTTGAAGCGCGCTGGGCCATG 20587  
CY 421 CAGTGCCTGTTGAGATGTCGAAACCGGACCGGTCATCCGGGAAATCTTTCGAGTACTC 480  
Db 20588 CAGTGCCTGTTGAGATGTCGAAACCGGACCGGTCATCCGGGAAATCTTTCGAGTACTC 20647  
CY 481 AAACCCGGTGGATCTCTCGCGTCAACGAGGTGTCMAACGAAACGAGCGCGCGAGT 540

Db 20648 AACCCTGCTGATCTCTGCGGCTCAACCGAGTCTGTCAACGAGAAAGCGGCGCGGAGT 20707  
Qy 541 CCGGTGTCCGGGAGACAGTGGCCGACCGGCTTCCGATTCGCTGCTAGCACTTCTG 600  
Db 20708 CCGGTGTCCGGGAGACAGTGGCCGACCGGCTTCCGATTCGCTGCTAGCACTTCTG 20767  
Qy 601 GAATGCTGCGGACAGCGGGGTTTCGATCTCGATTGGAGAGACGTCGTCGAGAGCC 660  
Db 20768 GAATGCTGCGGACAGCGGGGTTTCGATCTCGATTGGAGAGACGTCGTCGAGAGCC 20827  
Qy 661 CCGTACTTCATCCGACGATTCGCGGAGAGCTCGCTCGCACACAGCAGCGGATTCGGGAC 720  
Db 20828 CCGTACTTCATCCGACGATTCGCGGAGAGCTCGCTCGCACACAGCAGCGGATTCGGGAC 20887  
Qy 721 AGGTACGGGCGCGCTGTCCCGCTGGGCGCGCGCTGCTGCGATTATGAAATATGCC 780  
Db 20888 AGGTACGGGCGCGCTGTCCCGCTGGGCGCGCGCTGCTGCGATTATGAAATATGCC 20947  
Qy 781 CACGACATGGGCTATGCGATTCGACGCGGCGGAGCGGATTCGCTGA 828  
Db 20948 CACGACATGGGCTATGCGATTCGACGCGGCGGAGCGGATTCGCTGA 20995

RESULT 8  
AX600587/c 36538 bp DNA linear PAT 14-FEB-2003  
LOCUS AX600587  
DEFINITION Sequence 2 from Patent WO02079477.  
ACCESSION AX600587  
VERSION AX600587.1 GI:28400303  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS Hahn, D.R., Jackson, J.D., Bullard, B.S., Gustafson, G.D., Waldron, C. and Mitchell, J.C.  
TITLE Biosynthetic genes for butenyl-epinosyn insecticide production  
JOURNAL Patent: WO 02079477-A 2 10-OCT-2002;  
Dow Agrosciences LLC (US)  
FEATURES  
source 1. 36538  
/organism="Saccharopolyspora sp. NRRL 30141"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:221471"  
/note="LW107129"

ORIGIN  
Query Match 6.9%; Score 57; DB 6; Length 36538;  
Best Local Similarity 100.0%; Pred. No. 4,6e-18;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 73 TCGGTGCGGGCGCGCCCTGCGCATTCACACGCGCTACTGAGAGACGCGGCGG 129  
Db 866 TCGGTGCGGGCGCGCCCTGCGCATTCACACGCGCTACTGAGAGACGCGGCGG 810

RESULT 9  
AX600621/c 22 bp DNA linear PAT 14-FEB-2003  
LOCUS AX600621  
DEFINITION Sequence 36 from Patent WO02079477.  
ACCESSION AX600621  
VERSION AX600621.1 GI:28400307  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS Hahn, D.R., Jackson, J.D., Bullard, B.S., Gustafson, G.D., Waldron, C. and Mitchell, J.C.  
TITLE Biosynthetic genes for butenyl-epinosyn insecticide production  
JOURNAL Patent: WO 02079477-A 36 10-OCT-2002;

FEATURES  
source 1. 22  
/organism="Saccharopolyspora spinosa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:60894"

ORIGIN  
Query Match 2.7%; Score 22; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 769 GAGAAATATGCCACGACATGG 790  
Db 22 GAGAAATATGCCACGACATGG 1

RESULT 10  
E07847/c 457 bp DNA linear PAT 29-SEP-1997  
LOCUS E07847  
DEFINITION DNA sequence participating the biosynthesis of avermectin.  
ACCESSION E07847  
VERSION E07847.1 GI:2175980  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS Omura, S. and Ikeda, H.  
TITLE DIRECT FERMENTATION OF 5-OKO-AVERMECTIN DERIVATIVE AND MICROORGANISM FOR ITS PRODUCTION  
JOURNAL Patent: JP 1994189774-A 2 12-JUL-1994;  
KILASATO INST.THE  
COMMENT  
OS Streptomyces avermitilis  
PN JP 1994189774-A/2  
PD 12-JUL-1994  
PF 24-DEC-1992 JP 1992344605  
PI OMURA SATOSHI, IKEDA HARUO  
PC C12N15/53, C12N1/21, C12P19/62, (C12N1/21, C12R1:465), (C12P19/62, C12R1:465);  
PC C12R1:465);  
CC strandedness: Double;  
CC topology: Linear;  
FH key  
FH Location/Qualifiers  
FT source 1. 457  
/strain="K2062".  
/organism="Streptomyces avermitilis" FT  
/mol\_type="genomic DNA"  
/db\_xref="taxon:33903"

FEATURES  
source 1. 457  
/organism="Streptomyces avermitilis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:33903"

ORIGIN  
Query Match 2.7%; Score 22; DB 6; Length 457;  
Best Local Similarity 100.0%; Pred. No. 9.3;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 143 AGGCGCGGACCGGCTCACCGA 164  
Db 94 AGGCGCGGACCGGCTCACCGA 73

RESULT 11  
E07848/c 457 bp DNA linear PAT 29-SEP-1997  
LOCUS E07848  
DEFINITION DNA sequence participating the biosynthesis of avermectin.  
ACCESSION E07848  
VERSION E07848.1 GI:2175981  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS Omura, S. and Ikeda, H.  
TITLE DIRECT FERMENTATION OF 5-OKO-AVERMECTIN DERIVATIVE AND MICROORGANISM FOR ITS PRODUCTION  
JOURNAL Patent: JP 1994189774-A/3.  
Streptomyces avermitilis  
Streptomyces avermitilis

REFERENCE Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 457)  
AUTHORS Omura,S. and Ikeda,H.  
TITLE DIRECT FERMENTATION OF 5-OXO-AVERMECTIN DERIVATIVE AND MICROORGANISM FOR ITS PRODUCTION  
JOURNAL Patent: JP 1994189774-A 3 12-JUL-1994;  
KITASATO INST.THE  
COMMENT OS Streptomyces avermilitilis  
PN JP 1994189774-A/3  
PD 12-JUL-1994  
PF 24-DEC-1992 JP 1992344605  
PI OMURA SATOSHI, IKEDA HARUO  
PC C12N15/53,C12N1/21,C12P19/62,(C12N1/21,C12R1:465),(C12P19/62,C12R1:465);  
CC strandedness: Double;  
CC topology: Linear;  
FH Key Location/Qualifiers  
FT source 1..457  
/strain='K2082'.  
FEATURES location/Qualifiers  
source 1..457  
/organism='Streptomyces avermilitilis' FT  
1..457  
/mol\_type='genomic DNA'  
/db\_xref='taxon:33903'  
ORIGIN  
Query Match 2.7%; Score 22; DB 6; Length 457;  
Best Local Similarity 100.0%; Pred. No. 9.3;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 143 AGCGCCGACCGGCTCACCGA 164  
|||||  
94 AGCGCCGACCGGCTCACCGA 73  
DB  
RESULT 12  
E07846/c 1410 bp DNA linear PAT 29-SEP-1997  
LOCUS E07846  
DEFINITION DNA sequence participating the biosynthesis of avermectin.  
ACCESSION E07846  
VERSION E07846.1 GI:2175979  
KEYWORDS JP 1994189774-A/1.  
SOURCE Streptomyces avermilitilis  
ORGANISM Streptomyces avermilitilis  
Bacteria; Actinobacterii; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 1410)  
AUTHORS Omura,S. and Ikeda,H.  
TITLE DIRECT FERMENTATION OF 5-OXO-AVERMECTIN DERIVATIVE AND MICROORGANISM FOR ITS PRODUCTION  
JOURNAL Patent: JP 1994189774-A 1 12-JUL-1994;  
KITASATO INST.THE  
COMMENT OS Streptomyces avermilitilis  
PN JP 1994189774-A/1  
PD 12-JUL-1994  
PF 24-DEC-1992 JP 1992344605  
PI OMURA SATOSHI, IKEDA HARUO  
PC C12N15/53,C12N1/21,C12P19/62,(C12N1/21,C12R1:465),(C12P19/62,C12R1:465);  
CC strandedness: Double;  
CC topology: Linear;  
FH Key Location/Qualifiers  
FT source 1..1410  
/organism='Streptomyces avermilitilis' FT  
/strain='ATCC31271'.  
FEATURES location/Qualifiers  
source 1..1410  
/organism='Streptomyces avermilitilis'  
/mol\_type='genomic DNA'

ORIGIN /db\_xref='taxon:33903'  
Query Match 2.7%; Score 22; DB 6; Length 1410;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 143 AGCGCCGACCGGCTCACCGA 164  
|||||  
720 AGCGCCGACCGGCTCACCGA 699  
DB  
RESULT 13  
AB032524/c 5998 bp DNA linear BCT 23-SEP-1999  
LOCUS AB032524  
DEFINITION Streptomyces avermilitilis avermectin biosynthetic gene cluster (aver, aveF, aveD), complete cds.  
ACCESSION AB032524  
VERSION AB032524.1 GI:5921164  
KEYWORDS C5-O-methyltransferase, C-5 ketoreductase.  
SOURCE Streptomyces avermilitilis  
ORGANISM Streptomyces avermilitilis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 5998)  
AUTHORS Ikeda,H., Nonomura,T., Usami,M., Ohta,T. and Omura,S.  
TITLE Organization of the biosynthetic gene cluster for the polyketide antihelminthic macroide avermectin in Streptomyces avermilitilis  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)  
MEDLINE 99380548  
PUBMED 10449723  
REFERENCE 2 (bases 1 to 5998)  
AUTHORS Ikeda,H., Nonomura,T., Usami,M., Ohta,T. and Omura,S.  
TITLE Direct Submision  
JOURNAL Submitted (20-SEP-1999) Haruo Ikeda, School of Pharmaceutical Sciences, Kitasato University, Microbial Chemistry, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan  
(E-mail:ikedamc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242, Fax:+81-3-3444-6197)  
FEATURES location/Qualifiers  
source 1..5998  
/organism='Streptomyces avermilitilis'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:33903'  
gene 327..3176  
/gene='aver'  
327..3176  
/gene='aver'  
/note='contains helix-turn-helix DNA binding domain positive response regulator'  
/codon\_start=1  
/transl\_table=1  
/protein\_id='BA84600.1'  
/db\_xref='GI:5921165'  
/translation='WQGVSCIHPPKREBELTVDFETQPALRLATECAAGTYKLV  
AEGMGCGKSTFLGEALHTAASGFALVRAAGLPADRRQPLGVQQQLNDPAPEDTAR  
TAVPMVQVHVRGALERLAAGAPLAIIDVOADPSHCLKMTKTSPTSRLLIC  
TALACSPADPVLAEALMROTAFFERITLDCSLDGVGLSDRGQPTAALMALPVGNG  
TSEAARAARDDPADRPDADKPLPDSGAPQPPRDEGQCTTAALMADYCLT  
VTGGNPLLRALLLEHSEADAPASRAESALHSPOQAPPPVGVGRYGVSLATL  
SRRTTAIRQTNAGLAVUGGRADPLLEQLGASASTTRGLRLAEAGLITSGRRBP  
VAERAAIDALDPSSRAHLRRRAALQHHDAAPDVARRHAPRAAGPMAVSVLRBA  
AEOSLADDDVAVSACLEAVGACVRRERPEIRIRLAAAFGRNTNVAEHLADIVA  
TLREGEITGHQTLALVPLVNHGRLGARBARMLNADADRGICADGGEPPMAWPS  
TAHLAARDDPAARDPGRDPADKPLPDSGAPQPPRDEGQCTTAALMADYCLT  
TSEAARAARDDPADRPDADKPLPDSGAPQPPRDEGQCTTAALMADYCLT  
PQWQHHLAVRAELSTCRGLADAKCEQALTHVPHSHSVAPAGSLAOCVLACTAM  
GRIDEATQULSHVPEALPHSVGLGLRAGHFLMKNLPAVRRGTLTRAGVAREW  
GLDHPVLLPWRTPAAENFLRIGETKRADQQLITBOLVPHSGNPVVRGTALRLQOTA  
PAERILRLSVAVDLQSSGDRLLARALADLGAAYHSRNEPVRASATVRAMQJAKEC  
GAQALCDSLPSRGTDRGPDGRALAAARLAEALISEEMVATLAAGNTNREIAGRLCVT  
VSTVEQHLTVVRRKNTTRRELPTRLRHLDQNN'  
complement (3223..4131)  
/gene='aveF'  
gene



CDS  
complement(3223..4131)  
/gene="aveD"  
/note="post-polyketide modification"  
/codon\_start=1  
/transl\_table=11  
/product="C-5 ketoreductase"  
/protein\_id="BAA84601.1"  
/db\_xref="GI:5921166"  
translation="MRECGRGPVAVTRKVESLITSGPCLLSEKGNFVADLKRAA  
LVTGASRGIGRAIAQRLAEGLVAVHVSQSKAAETVEALERVGEPFVADLR  
DAVDLTFEVENRLDGRPLVAVNAGVSPAPGAGKGEVTSVGVAAEDH  
LPIINVAPEFIIOIALRLIGEGRVINVSALTIRIAMPLIPYAMSKALEMIAEFLA  
LEIGARGITVNTVAPGITATDNNRAMRDVPGVEERI.SALTALRI.GRPEDIGVAAFL  
ASDEGRWVTGQLVDSAGGLCTAPAMS"  
complement(4911..5762)  
/gene="aveD"  
complement(4911..5762)  
/gene="aveD"  
/note="post-polyketide modification"  
/codon\_start=1  
/transl\_table=11  
/product="C5-O-methyltransferase"  
/protein\_id="BAA84602.1"  
/db\_xref="GI:5921167"  
translation="MMAVPEQTPPSLSEVDYVDRITDLMNRALGNTHTLGYWPOGD  
GSPGKAADRLTDLIGKRGITRRVLDVGGSGGPAYRLASAPVDVGVTVSVQ  
VELATLAKQSHVADRVFTRADAMELPPDPSFDAMLECLHWPSPAQVTEAR  
VLRPGRLAVTIVALLRAFGRTKRGECSTQLAVPALVHIDVYAGMIDAGLEIHEL  
TDIGDOVGPSPALRDHYNEHLDEYAAAFGIVAMRKVAQCITLTPWTPIGYVVL  
TARRPGE"  
ORIGIN  
Query Match 2.7%; Score 22; DB 1; Length 5998;  
Best Local Similarity 100.0%; Pred. No. 8.4;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 143 AGCGCCGCGACCGGCTCACC GA 164  
DB 5614 AGCGCCGCGACCGGCTCACC GA 5593  
RESULT 14  
BD217469/c 12381 bp DNA linear PAT 17-JUL-2003  
LOCUS Polyketides, their preparation, and materials for use therein.  
ACCESSION BD217469  
VERSION BD217469.1 GI:33027239  
KEYWORDS JP 2002519066-A/1  
SOURCE Streptomyces avermitilis  
ORGANISM Streptomyces avermitilis  
Bacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
Kellenberger, J.L., Leadlay, P.F., Staunton, J., Engwall, K.J.S. and  
McArthur, H.A.1.  
Polyketides, their preparation, and materials for use therein  
Patent: JP 2002519066-A 1 02-JUL-2002;  
BIOTICA TECHNOLOGY LTD, PFIZER INC  
OS Streptomyces avermitilis  
PN JP 2002519066-A/1  
PD 02-JUL-2002 JP 2000558217  
PF 06-JUL-1999 JP 2000558217  
PI 06-JUL-1998 GB 9814622.8  
PI JOHANNES LAURENZ KELLENBERGER, PETER FRANCIS LEADLAY, JAMES PI  
STAUNTON,  
PT KIM JONELLE STUTZMAN ENGMALL, HAMISH ALASTAIR IRVINE MCARTHUR  
PC C12N15/09, C07K19/00, C12N1/21, C12N9/02, C12N9/04, C12P17/06, PC  
C12P19/62//  
PC (C12N1/21, C12R1:465), C12N15/00  
CC Polyketides, their preparation, and materials for use therein  
FH Key Location/Qualifiers  
FT source 1..12381  
FT /organism="Streptomyces avermitilis".

FEATURES  
source Location/Qualifiers  
1..12381  
/organism="Streptomyces avermitilis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:33903"  
ORIGIN  
Query Match 2.7%; Score 22; DB 6; Length 12381;  
Best Local Similarity 100.0%; Pred. No. 8.1;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 143 AGCGCCGCGACCGGCTCACC GA 164  
DB 333 AGCGCCGCGACCGGCTCACC GA 312  
RESULT 15  
AX006889/c 12381 bp DNA linear PAT 06-SEP-2000  
LOCUS AX006889  
DEFINITION Sequence 1 from Patent WO0001827.  
ACCESSION AX006889  
VERSION AX006889.1 GI:9994904  
KEYWORDS Streptomyces avermitilis  
SOURCE Streptomyces avermitilis  
ORGANISM Streptomyces avermitilis  
Bacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1  
Kellenberger, J.L., Leadlay, P.F., Staunton, J., McArthur, H.A. and  
Stutzman-Engwall, K.J.  
Polyketides, their preparation, and materials for use therein  
Patent: WO 0001827-A 1 13-JAN-2000;  
KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);  
STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); MCARTHUR  
HAMISH ALASTAIR IRVINE (US); STUTZMAN ENGMALL KIM JONELLE (US)  
FEATURES  
source Location/Qualifiers  
1..12381  
/organism="Streptomyces avermitilis"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:33903"  
ORIGIN  
Query Match 2.7%; Score 22; DB 6; Length 12381;  
Best Local Similarity 100.0%; Pred. No. 8.1;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 143 AGCGCCGCGACCGGCTCACC GA 164  
DB 333 AGCGCCGCGACCGGCTCACC GA 312  
Search completed: August 29, 2005, 18:38:21  
Job time : 3897 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 29, 2005, 15:06:10 ; Search time 585 Seconds  
(without alignments)  
8378.711 Million cell updates/sec

Title: US-10-069-353a-7

Sequence: 1 ggtgtgcacagtggtgcacc.....cgcggaagccggtcggtcga 828

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4390206 seqs, 2959870667 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	828	100.0	828	4 AAF88318	AAf88318 S. spinos
2	828	100.0	45624	4 AAF88315	AAf88315 S. spinos
3	828	100.0	50000	4 AAF88312	AAf88312 S. spinos
4	828	100.0	80161	2 AA221501	AA221501 DNA fragm
5	57	6.9	36538	10 ABV75558	ABv75558 Saccharop
6	22	2.7	22	10 ABV75562	ABv75562 Saccharop
7	22	2.7	457	2 AAQ70316	AAq70316 Oxidoredu
8	22	2.7	457	2 AAQ70315	AAq70315 Oxidoredu
9	22	2.7	1410	2 AAQ80524	AAq80524 Oxidoredu
10	22	2.7	12381	3 AA258381	AA258381 Streptomy
11	20	2.4	20	10 ABV75561	ABv75561 Saccharop
12	20	2.4	237	12 ADP62503	ADp62503 Maize car
13	20	2.4	263	12 ADP62374	ADp62374 Maize car
14	20	2.4	290	12 ADP62375	ADp62375 Maize car
15	20	2.4	294	12 ADP62372	ADp62372 Maize car
16	20	2.4	305	12 ADP62372	ADp62372 Maize car
17	20	2.4	433	13 ADP62502	ADp62502 Maize car
18	20	2.4	1164	13 ADP62502	ADp62502 Maize car
19	20	2.4	1650	3 AAC44549	AAC44549 Zea mays
20	20	2.4	1832	6 AAL45918	AAL45918 Terrabact

#### ALIGNMENTS

RESULT 1	21	20	2.4	2805	8	ACA40862
AAf88318	22	20	2.4	2985	8	ACA38551
AAf88318 standard; DNA; 828 BP.	23	20	2.4	3107	2	AAV68092
AAf88318;	24	20	2.4	8051	2	AAf72685
28-AUG-2001 (first entry)	25	20	2.4	8160	2	AAx25774
S. spinosa DNA fragment encoding ORF1, SEQ ID 7.	26	20	2.4	110000	4	AAI99682_40
Forosamine; trimethylthamose; polyketide synthase; biosynthesis;	27	20	2.4	110000	4	AAI99683_40
spinosyn; polyketide aglycone; transgenic plant; insect resistance;	28	19	2.3	909	5	AAH94478
macrolide; insecticidal; O-methyltransferase; ds.	29	19	2.3	1050	5	AAH94372
Saccharopolyspora spinosa.	30	19	2.3	1188	10	ADP36141
OS	31	19	2.3	1418	10	ADP36144
XX	32	19	2.3	1418	10	ADP36202
XX	33	19	2.3	4741	4	ABL27984
XX	34	19	2.3	110000	2	AAx91990_03
XX	35	19	2.3	110000	2	AAx91990_04
XX	36	19	2.3	273254	3	AAc81914
XX	37	18	2.2	266	10	ABX84638
XX	38	18	2.2	268	7	ADP70438
XX	39	18	2.2	272	10	ABX83984
XX	40	18	2.2	273	7	ADP70078
XX	41	18	2.2	275	7	ADP70280
XX	42	18	2.2	276	7	ADP70495
XX	43	18	2.2	276	7	ADP70414
XX	44	18	2.2	285	7	ADP70463
XX	45	18	2.2	295	6	ABL73622
ACA40862 Prokaryot						
ACA38551 Prokaryot						
AAV68092 cDNA enco						
AAf72685 Sugar bio						
AAx25774 S. erythra						
AAI99682_40 Continuation (41 o						
AAI99683_40 Continuation (41 o						
AAH94478 Human foe						
AAH94372 Human foe						
ADP36141 Weed cont						
ADP36144 Weed cont						
ADP36202 Weed cont						
ABL27984 Drosophila						
Continuation (4 of						
Continuation (5 of						
AAc81914 Chlamydia						
ABX84638 Corn ear-						
ADP70438 Corn seed						
ABX83984 Corn ear-						
ADP70078 Corn seed						
ADP70280 Corn seed						
ADP70495 Corn seed						
ADP70414 Corn seed						
ADP70463 Corn seed						
ABL73622 Corn tass						

Claim 7a; Page 102-104; 354pp; German.

This invention describes a novel method nucleic acid (I) and its encoded polypeptide (II) containing at least one region that encodes an enzymatic activity involved in biosynthesis of spinosyns. (I) are used (i) to identify, inactivate or modulate genes involved in the biosynthesis of (II); (ii) to generate a library of polyketide synthases; (iii) for adding forosamine or trimethylthamose to a spinosyn or polyketide aglycone; and (iv) for recombinant production of the corresponding

enzymes, which are used for production of (II), their precursors or derivatives, including production of transgenic plants that express (II) and thus have increased resistance to insects. (I) are also useful as markers for sequencing of the Saccharopolyspora spinosa genome. (II) are macrolides with insecticidal, but not antibacterial, activity, and can also be used to raise specific antibodies, useful for identifying expression clones in a gene bank. Cells transformed with (I) may produce (II) at significantly increased levels or produce new derivatives of (II). This sequence, ORF 1, encodes an S. spinosa O-methyltransferase protein

XX Sequence 828 BP; 140 A; 262 C; 285 G; 141 T; 0 U; 0 Other;

Query Match 100.0%; Score 828; DB 4; Length 828;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTTGCCAGTGGCGCACCAACATCGACAGTTGGGAGATGATGACTGTGACG 60  
 DB 1 GTGTTGCCAGTGGCGCACCAACATCGACAGTTGGGAGATGATGACTGTGACG 60

QY 61 CCGTTGCTGAACCTGGTGGGGGGCCCTGGCCATCAGCGGCTACTGGGAGAAC 120  
 DB 61 CCGTTGCTGAACCTGGTGGGGGGCCCTGGCCATCAGCGGCTACTGGGAGAAC 120

QY 121 GACGGGCGGGCTTCTGCGACAGAGCGCGGCTGACCGACTTGTCCGCGAACGG 180  
 DB 121 GACGGGCGGGCTTCTGCGACAGAGCGCGGCTGACCGACTTGTCCGCGAACGG 180

QY 181 ACCGTTCTGATGAGCGGCGCTTGTGATGTTGGGGTGGGTTACCGGACACCGAGC 240  
 DB 181 ACCGTTCTGATGAGCGGCGCTTGTGATGTTGGGGTGGGTTACCGGACACCGAGC 240

QY 241 CCGCGGCTGGCGGCGACGACGATCCAGATCCCGGATCAGCGGTGACGAGGTGCA 300  
 DB 241 CCGCGGCTGGCGGCGACGACGATCCAGATCCCGGATCAGCGGTGACGAGGTGCA 300

QY 301 GTGGCCATGCGCGCTGATTCGCGACGCGACCGGACTAAGCCACCGGCTGACTTCTG 360  
 DB 301 GTGGCCATGCGCGCTGATTCGCGACGCGACCGGACTAAGCCACCGGCTGACTTCTG 360

QY 361 TGGCGTGAATGCATGCTCCCTGCGGTTACCGGACAACTGTTTGACGCGCGCTGGCCATG 420  
 DB 361 TGGCGTGAATGCATGCTCCCTGCGGTTACCGGACAACTGTTTGACGCGCGCTGGCCATG 420

QY 421 CAGTGGCTTGGAGATGTCGGAACCGGACCGGTCATCCGGGAAATCTTGAGATCTC 480  
 DB 421 CAGTGGCTTGGAGATGTCGGAACCGGACCGGTCATCCGGGAAATCTTGAGATCTC 480

QY 481 AAACCCGGTGGCATCTCGGCGTCAACGAGTGTCAAAAGAAACGGGCGCGGAGTG 540  
 DB 481 AAACCCGGTGGCATCTCGGCGTCAACGAGTGTCAAAAGAAACGGGCGCGGAGTG 540

QY 541 CCGGTGTCCGGGAGACAGGTGGCGGACCGGCTTGGGATGCTGCTGGCTGAGAACTTCTG 600  
 DB 541 CCGGTGTCCGGGAGACAGGTGGCGGACCGGCTTGGGATGCTGCTGGCTGAGAACTTCTG 600

QY 601 GAATGCTGCTGACAGCGGGGTTTCGAGATCTCGATTGGGAGAGAGTGTCGCGAGAAC 660  
 DB 601 GAATGCTGCTGACAGCGGGGTTTCGAGATCTCGATTGGGAGAGAGTGTCGCGAGAAC 660

QY 661 CCGTACTTCATGCGCGAGTTTCGCGAAGAGTCTGCTGGGACACAGACGGAGATCCGGAC 720  
 DB 661 CCGTACTTCATGCGCGAGTTTCGCGAAGAGTCTGCTGGGACACAGACGGAGATCCGGAC 720

QY 721 AGGTACGGGGCCGGCTGTGCGCGCTGGGCGCGCGGCTGCGATTATGAGAAATATGCC 780  
 DB 721 AGGTACGGGGCCGGCTGTGCGCGCTGGGCGCGCGGCTGCGATTATGAGAAATATGCC 780

QY 781 CACGACATGGGCTATGCGATCTGACAGGCGCGGAAGCGGCTGAGCTGA 828  
 DB 781 CACGACATGGGCTATGCGATCTGACAGGCGCGGAAGCGGCTGAGCTGA 828

# RESULT 2

AAF88315/C  
 ID AAF88315 standard; DNA; 45624 BP.

AAF88315;

28-AUG-2001 (first entry)

S. spinosa DNA fragment SEQ ID 4.

Forosamine; trimethylrhamsone; polyketide synthase; biosynthesis;

spinosyn; polyketide aglycone; transgenic plant; insect resistance;

macrolide; insecticidal; ds.

Saccharopolyspora spinosa.

DE19957268-A1.

08-MAR-2001.

29-NOV-1999; 99DE-01057268.

27-AUG-1999; 99DE-01040596.

(FARB ) BAYER AG.

Eberz G, Moehrl V, Froede R, Velten R, Salas JA;

WPI; 2001-267102/28.

New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for recombinant production of insecticidal spinosyns and their derivatives.

Claim 7; Page 56-74; 354pp; German.

This invention describes a novel method nucleic acid (I) and its encoded polypeptide (II) containing at least one region that encodes an enzymatic activity involved in biosynthesis of spinosyns. (I) are used (i) to identify, inactivate or modulate genes involved in the biosynthesis of (II); (ii) to generate a library of polyketide synthases; (iii) for adding forosamine or trimethylrhamsone to a spinosyn or polyketide aglycone; and (iv) for recombinant production of the corresponding enzymes, which are used for production of (II), their precursors or derivatives, including production of transgenic plants that express (II) and thus have increased resistance to insects. (I) are also useful as markers for sequencing of the Saccharopolyspora spinosa genome. (II) are macrolides with insecticidal, but not antibacterial, activity, and can also be used to raise specific antibodies, useful for identifying expression clones in a gene bank. Cells transformed with (I) may produce (II) at significantly increased levels or produce new derivatives of (II). This sequence represents a genomic DNA fragment of the S. spinosa genome which contains the coding regions for proteins involved in forosamine and trimethylrhamsone biosynthesis

XX Sequence 45624 BP; 7933 A; 14913 C; 14940 G; 7838 T; 0 U; 0 Other;

Query Match 100.0%; Score 828; DB 4; Length 45624;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTTGCCAGTGGCGCACCAACATCGACAGTTGGGAGATGATGACTGTGACG 60  
 DB 828 GTGTTGCCAGTGGCGCACCAACATCGACAGTTGGGAGATGATGACTGTGACG 769

QY 61 CCGTTGCTGAACCTGGTGGGGGGCCCTGGCCATCCACCGGCTACTGGGAGAAC 120  
 DB 768 CCGTTGCTGAACCTGGTGGGGGGCCCTGGCCATCCACCGGCTACTGGGAGAAC 709

QY 121 GACGGGCGGGCTTCTGCGACGAGCGCGGCTGACCGGCTTGTGCGCGAACGG 180  
 DB 708 GACGGGCGGGCTTCTGCGACGAGCGCGGCTGACCGGCTTGTGCGCGAACGG 649

```

QY 181 ACCGCTGCTGATGAGGCGGCTTGCATGCTGATGAGGAGTGGGATGCGCAACCGACG 240
DB 648 ACCGCTGCTGATGAGGCGGCTTGCATGCTGATGAGGAGTGGGATGCGCAACCGACG 589
QY 241 CTGCGCGTGCAGCGGACGACGATCCAGATCACCAGGATCACCAGGATGCA 300
DB 588 CTGCGCGTGCAGCGGACGACGATCCAGATCACCAGGATCACCAGGATGCA 529
QY 301 GTGCGCATGCGCGTGCATGAGCGACGCGACGCGGATCAGGATGCACTTCTG 360
DB 528 GTGCGCATGCGCGTGCATGAGCGACGCGACGCGGATCAGGATGCACTTCTG 469
QY 361 TGGCGTATGCGCATGCTCCGCGGACCGGATCAGGATGCACTTCTG 420
DB 468 TGGCGTATGCGCATGCTCCGCGGACCGGATCAGGATGCACTTCTG 409
QY 421 CAGTGCCTGTTGAGATGTCGACCGGACCGGATCAGGATGCACTTCTG 480
DB 408 CAGTGCCTGTTGAGATGTCGACCGGACCGGATCAGGATGCACTTCTG 349
QY 481 AAACCCGGTGGCATCTCGCGCTGACCGGATGTCGACCGGATGCACTTCTG 540
DB 348 AAACCCGGTGGCATCTCGCGCTGACCGGATGTCGACCGGATGCACTTCTG 289
QY 541 CCGGCTGCTCGGAGGACAGGATGCGGACCGGATGTCGACCGGATGCACTTCTG 600
DB 288 CCGGCTGCTCGGAGGACAGGATGCGGACCGGATGTCGACCGGATGCACTTCTG 229
QY 601 GAATGCTGCTGAGCGGAGGATGTCGACCGGATGTCGACCGGATGCACTTCTG 660
DB 228 GAATGCTGCTGAGCGGAGGATGTCGACCGGATGTCGACCGGATGCACTTCTG 169
QY 661 CCGTACTTCATGCGCGGATGTCGACCGGATGTCGACCGGATGCACTTCTG 720
DB 168 CCGTACTTCATGCGCGGATGTCGACCGGATGTCGACCGGATGCACTTCTG 109
QY 721 AGGTACCGGCGGCGGATGTCGACCGGATGTCGACCGGATGCACTTCTG 780
DB 108 AGGTACCGGCGGCGGATGTCGACCGGATGTCGACCGGATGCACTTCTG 49
QY 781 CACGACATGCGGATGTCGACCGGATGTCGACCGGATGCACTTCTG 828
DB 48 CACGACATGCGGATGTCGACCGGATGTCGACCGGATGCACTTCTG 1

```

RESULT 3  
AAF88312 strand; DNA; 50000 BP.

AAF88312;  
28-AUG-2001 (first entry)  
S. spinosa DNA fragment SEQ ID 1.  
Forosamine; trimethylrhinmose; polyketide synthase; biosynthesis;  
spinosyn; polyketide aglycone; transgenic plant; insect resistance;  
macrolide; insecticidal; ds.  
Saccharopolyspora spinosa.  
DE19957268-A1.  
08-MAR-2001.  
29-NOV-1999; 99DE-01057268.  
27-AUG-1999; 99DE-01040596.  
(FARB ) BAYER AG.  
Ebertz G, Moehle V, Froede R, Velten R, Salas JA;

```

DR WPI, 2001-267102/28.
XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives.
XX Claim 7; Page 14-31, 354pp; German.
XX This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (ii) to generate a library of polyketide synthases; (iii) for
CC adding forosamine or trimethylrhinmose to a spinosyn or polyketide
CC aglycone; and (iv) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence represents a genomic DNA fragment of the S. spinosa
CC genome which contains the coding regions for proteins involved in
CC forosamine, trimethylrhinmose and polyketide synthase biosynthesis
XX
XX Sequence 50000 BP; 8484 A; 16290 C; 16535 G; 8691 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 828; DB 4; Length 50000;
Beet Local Similarity 100.0%; Pred. No. 0;
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGTGCGACAGTGGCGGACCAACATCGAGAGGTTGGGCGAGATGTAAGCTGTGACG 60
DB 44797 GTGTGCGACAGTGGCGGACCAACATCGAGAGGTTGGGCGAGATGTAAGCTGTGACG 44856
QY 61 CCGTGTGCTGAACTCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
DB 44857 CCGTGTGCTGAACTCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 44916
QY 121 GACGCGGCGGCGGCTTCTGCGGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
DB 44917 GACGCGGCGGCGGCTTCTGCGGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 44976
QY 181 ACCGTGCTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
DB 44977 ACCGTGCTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 45036
QY 241 CTGCGCGTGCAGCGGACGACGATCCAGATCACCAGGATCACCAGGATGCA 300
DB 45037 CTGCGCGTGCAGCGGACGACGATCCAGATCACCAGGATCACCAGGATGCA 45096
QY 301 GTGCGCATGCGCGTGCATGAGCGACGCGGATCAGGATGCACTTCTG 360
DB 45097 GTGCGCATGCGCGTGCATGAGCGACGCGGATCAGGATGCACTTCTG 45156
QY 361 TGGCGTATGCGCATGCTCCGCGGACCGGATCAGGATGCACTTCTG 420
DB 45157 TGGCGTATGCGCATGCTCCGCGGACCGGATCAGGATGCACTTCTG 45216
QY 421 CAGTGCCTGTTGAGATGTCGACCGGACCGGATCAGGATGCACTTCTG 480
DB 45217 CAGTGCCTGTTGAGATGTCGACCGGACCGGATCAGGATGCACTTCTG 45276
QY 481 AAACCCGGTGGCATCTCGCGCTGACCGGATGTCGACCGGATGCACTTCTG 540
DB 45277 AAACCCGGTGGCATCTCGCGCTGACCGGATGTCGACCGGATGCACTTCTG 45336
QY 541 CCGGCTGCTCGGAGGACAGGATGCGGACCGGATGTCGACCGGATGCACTTCTG 600
DB 45337 CCGGCTGCTCGGAGGACAGGATGCGGACCGGATGTCGACCGGATGCACTTCTG 45396
QY 601 GAATGCTGCTGAGCGGAGGATGTCGACCGGATGTCGACCGGATGCACTTCTG 660

```

Db 45397 GAATGCTGCGTGAGCGGGGTTGAGATCCTCGATTGGAGAGGTGTCGTGAGGACC 45456  
Oy 661 CGGTACTTCATGCCGCACTTGGCCGAAGAAGCTCGTGCGACCAAGCAGGATCGCGGAC 720  
Db 45457 CGGTACTTCATGCCGCACTTGGCCGAAGAAGCTCGTGCGACCAAGCAGGATCGCGGAC 45516  
Oy 721 AGGTACGGGCGGCGGTGTCGCGCGCTGGCGCCGCGCGGTCTCGATTATGAGAAATATGCG 780  
Db 45517 AGGTACGGGCGGCGGTGTCGCGCGCTGGCGCCGCGCGGTCTCGATTATGAGAAATATGCC 45576  
Oy 781 CACGACATGGGCTATGCGATTCTGACGGCGCGGAAAGCGGTGCGCTGA 828  
Db 45577 CACGACATGGGCTATGCGATTCTGACGGCGCGGAAAGCGGTGCGCTGA 45624

RESULT 4  
AA221501  
ID AA221501 standard; DNA; 80161 BP.  
XX  
XX AA221501;  
XX  
XX 01-DEC-1999 (first entry)  
XX  
DE DNA fragment of Saccharopolyspora spinosa containing biosynthetic genes.  
XX  
XX Spinosyn biosynthetic enzyme; open reading frame; ORF.  
KM insecticidal microlides; arachnid; nematode; insect; polyketide;  
KM polyketide synthase; PKS; extender module; initiator module;  
KM acyl transferase domain; AT; acyl carrier protein; ACP;  
KM beta-ketosynthase domain; KS; KR; dehydratase domain; DH;  
KM enoyl reductase domain; ER; beta-ketoreductase; insecticide; ss.  
XX  
OS Saccharopolyspora spinosa.  
XX  
XX  
FH Key Location/Qualifiers  
FT complement(1135. .1971)  
FT /\*tag= a  
FT /product= "ORFL16"  
FT /note= "Protein involved in transcription control"  
FT 2024. .2791  
FT /\*tag= b  
FT /product= "ORFL15"  
FT /note= "Keto acyl reductase"  
FT complement(3416. .4165)  
FT /\*tag= c  
FT /product= "spnS"  
FT /note= "Spinosyn biosynthesis protein S"  
FT complement(4168. .5325)  
FT /\*tag= d  
FT /product= "spnR"  
FT /note= "Spinosyn biosynthesis protein R"  
FT complement(5363. .6751)  
FT /\*tag= e  
FT /product= "spnQ"  
FT /note= "Spinosyn biosynthesis protein Q"  
FT 7083. .8450  
FT /\*tag= f  
FT /product= "spnP"  
FT /note= "Spinosyn biosynthesis protein P"  
FT 8967. .10427  
FT /\*tag= g  
FT /product= "spnO"  
FT /note= "Spinosyn biosynthesis protein O"  
FT 10436. .11434  
FT /\*tag= h  
FT /product= "spnN"  
FT /note= "Spinosyn biosynthesis protein N"  
FT complement(11530. .12492)  
FT /\*tag= i  
FT /product= "spnM"  
FT /note= "Spinosyn biosynthesis protein M"  
FT complement(12696. .13547)  
FT CDS

FT /\*tag= j  
FT /product= "spnL"  
FT /note= "Spinosyn biosynthesis protein L"  
FT complement(13592. .14785)  
FT /\*tag= k  
FT /product= "spnK"  
FT /note= "Spinosyn biosynthesis protein K"  
FT complement(14799. .16418)  
FT /\*tag= l  
FT /product= "spnJ"  
FT /note= "Spinosyn biosynthesis protein J"  
FT 16556. .17743  
FT /\*tag= m  
FT /product= "spnI"  
FT /note= "Spinosyn biosynthesis protein I"  
FT complement(17749. .18501)  
FT /\*tag= n  
FT /product= "spnH"  
FT /note= "Spinosyn biosynthesis protein H"  
FT complement(18541. .19713)  
FT /\*tag= o  
FT /product= "spnG"  
FT /note= "Spinosyn biosynthesis protein G"  
FT 20168. .20995  
FT /\*tag= p  
FT /product= "spnF"  
FT /note= "Spinosyn biosynthesis protein F"  
FT 21111. .28898  
FT /\*tag= q  
FT /product= "spnA"  
FT /note= "Spinosyn biosynthesis protein A"  
FT /transl\_except= (Pos:26940. .26942, aa:Pro)  
FT 21126. .22379  
FT /\*tag= r  
FT /note= "Beta-ketosynthase domain (KS1): part of the  
FT initiator module"  
FT 22692. .23669  
FT /\*tag= s  
FT /note= "Acyl transferase domain (AT1): part of the  
FT initiator module"  
FT 23793. .24041  
FT /\*tag= t  
FT /note= "Acyl carrier protein domain (ACPI): part of the  
FT initiator module"  
FT 24102. .25349  
FT /\*tag= u  
FT /note= "Beta-ketosynthase domain (KS1): part of extender  
FT module 1"  
FT 25683. .26684  
FT /\*tag= v  
FT /note= "Acyl transferase domain (AT1): part of extender  
FT module 1"  
FT 27582. .28121  
FT /\*tag= w  
FT /note= "Beta-ketoreductase domain (KR1): part of extender  
FT module 1"  
FT 28404. .28649  
FT /\*tag= x  
FT /note= "Acyl carrier protein domain (ACPI): part of  
FT extender module 1"  
FT 29024. .30295  
FT /\*tag= z  
FT /note= "Beta-ketosynthase domain (KS2): part of extender  
FT module 2"  
FT 30629. .31621  
FT /\*tag= aa  
FT /note= "Acyl transferase domain (AT2): part of extender  
FT module 2"  
FT 31697. .32254  
FT /\*tag= ab  
FT /note= "Dehydratase domain (DH2): part of extender module  
FT 2"  
FT 33035. .34072  
FT misc\_feature

```
FT      /*tag= ac
FT      /note= "Enoyl reductase domain (ER2) : part of extender
FT      module 2"
FT      34082. .34621
FT      /*tag= ad
FT      /note= "Beta-ketoreductase domain (KR2) : part of extender
FT      module 2"
FT      34886. .30295
FT      /*tag= ae
FT      /note= "Acyl carrier protein domain (ACP2) : part of
FT      extender module 2"
FT      35419. .44931
FT      /*tag= af
FT      /product= "spnc"
FT      /note= "Spinosyn biosynthesis protein C"
FT      35518. .36786
FT      /*tag= ag
FT      /note= "Beta-ketosynthase domain (KS3) : part of extender
FT      module 3"
FT      37108. .38097
FT      /*tag= ah
FT      /note= "Acyl transferase domain (AT3) : part of extender
FT      module 3"
FT      38916. .35374
FT      /*tag= y
FT      /product= "spnc"
FT      /note= "Spinosyn biosynthesis protein C"
FT      38992. .39528
FT      /*tag= ai
FT      /note= "Beta-ketoreductase domain (KR3) : part of extender
FT      module 3"
FT      39790. .40035
FT      /*tag= aj
FT      /note= "Acyl carrier protein domain (ACP3) : part of
FT      extender module 3"
FT      40102. .41373
FT      /*tag= ak
FT      /note= "Beta-ketosynthase domain (KS4) : part of extender
FT      module 4"
FT      41713. .42705
FT      /*tag= al
FT      /note= "Acyl transferase domain (AT4) : part of extender
FT      module 4"
FT      43615. .44157
FT      /*tag= am
FT      /note= "Beta-ketoreductase domain (KR4) : part of extender
FT      module 4"
FT      44431. .44676
FT      /*tag= an
FT      /note= "Acyl carrier protein domain (ACP4) : part of
FT      extender module 4"
FT      44966. .59752
FT      /*tag= ao
FT      /product= "spnd"
FT      /note= "Spinosyn biosynthesis protein D"
FT      45077. .46348
FT      /*tag= ap
FT      /note= "Beta-ketosynthase domain (KS5) : part of extender
FT      module 5"
FT      46691. .47674
FT      /*tag= aq
FT      /note= "Acyl transferase domain (AT5) : part of extender
FT      module 5"
FT      47753. .48310
FT      /*tag= ar
FT      /note= "Dehydratase domain (DHS) : part of extender module
FT      5"
FT      49226. .49771
FT      /*tag= as
FT      /note= "Beta-ketoreductase domain (KR5) : part of extender
FT      module 5"
FT      50009. .50254
FT      misc_feature
```

```
Query Match      100.0%; Score 828; DB 2; Length 80161;
Best Local Similarity 100.0%; Pred. No. 0; Indels 0; Gaps 0;
Matches 828; Conservative 0; Mismatches 0;

Qy      1      GTGTTCCAGATGCGCCACCAATCGCAGCAGGTGGCAGATGATGACCTGTACG 60
Db      20168  GTGTTCCAGATGCGCCACCAATCGCAGCAGGTGGCAGATGATGACCTGTACG 20227

Qy      61      CCGTGTCTGAATCTCGGTGCGCGGCGGCGCCCTGCGCCATCAACACGGCTACTGCGGAAAC 120
Db      20228  CCGTGTCTGAATCTCGGTGCGCGGCGGCGCCCTGCGCCATCAACACGGCTACTGCGGAAAC 20287

Qy      121     GACGGGCGGGCTTCCGTGGAGAGAGCGCGCAACCGGCTCAGCAGACCTGTGCGGAAACG 180
Db      20288  GACGGGCGGGCTTCCGTGGAGAGAGCGCGCAACCGGCTCAGCAGACCTGTGCGGAAACG 20347

Qy      181     ACCGTCTCGATGACCGGCGGTTCGACTGTGATGTTGGGGTGGCGTACCGGACCAACGAGC 240
Db      20348  ACCGTCTCGATGACCGGCGGTTCGACTGTGATGTTGGGGTGGCGTACCGGACCAACGAGC 20407

Qy      241     CTGCGCGTGGCGCGCGCAACCGGATCCAGATCAACCGCATCACCGTCAAGCAGGTGCA 300
Db      20408  CTGCGCGTGGCGCGCGCAACCGGATCCAGATCAACCGCATCACCGTCAAGCAGGTGCA 20467

Qy      301     GTGGCCATGCGCGGTGATGGCGGAGCGGAAACGGGAGCTAAAGCACCGGGTGAATTCTCG 360
Db      20468  GTGGCCATGCGCGGTGATGGCGGAGCGGAAACGGGAGCTAAAGCACCGGGTGAATTCTCG 20527

Qy      361     TGCGTCGATGCAATGTCCTGCGGTACCGGCAATCTTTCGACGCGCGCTGGGCGATG 420
Db      20528  TGCGTCGATGCAATGTCCTGCGGTACCGGCAATCTTTCGACGCGCGCTGGGCGATG 20587

Qy      421     CAGTCGCTGTTGGAAGATGTCGCAACCGGACCGTGGCATTCGGGAAATCTTTCAGATATC 480
Db      20588  CAGTCGCTGTTGGAAGATGTCGCAACCGGACCGTGGCATTCGGGAAATCTTTCAGATATC 20647

Qy      481     AAACCCGGTGGCATCTCCGCGGTGCAACGAGTGTGTAACGAGAGCGGCGGAGATG 540
Db      20648  AAACCCGGTGGCATCTCCGCGGTGCAACGAGTGTGTAACGAGAGCGGCGGAGATG 20707

Qy      541     CCGGTGTCGGGAGCAGATGCGCGGCGGCTTCGATCTGCTGCTGAGCAATCTTCTG 600
Db      20708  CCGGTGTCGGGAGCAGATGCGCGGCGGCTTCGATCTGCTGCTGAGCAATCTTCTG 20767

Qy      601     GAATCGCTGCTGTCAGCGGGGTTTCAGATCTCTGATTTGGAGAGACGTTGTGAGAGACC 660
Db      20768  GAATCGCTGCTGTCAGCGGGGTTTCAGATCTCTGATTTGGAGAGACGTTGTGAGAGACC 20827

Qy      661     CGGTACTTCAATGCGCGAGTTCCGCGAAGCTGCTGCGGACACGACCGGATCGCGGAC 720
Db      20828  CGGTACTTCAATGCGCGAGTTCCGCGAAGCTGCTGCGGACACGACCGGATCGCGGAC 20887

Qy      721     AGGTACGGGCGGCGCTGTCGCGGCTGCGGCTGCGGCTGCTGATTAAGAAATATGCC 780
Db      20888  AGGTACGGGCGGCGCTGTCGCGGCTGCGGCTGCGGCTGCTGATTAAGAAATATGCC 20947

Qy      781     CACGACATGGGCTATGCGATTCTGACGCGCGCGGAAGCCGCTCGGCTGCA 828
Db      20948  CACGACATGGGCTATGCGATTCTGACGCGCGCGGAAGCCGCTCGGCTGCA 20995

RESULT 5
ABV75558/c
ID      ABV75558 standard; DNA; 36538 BP.
XX      AC      ABV75558;
XX      AC      22-JAN-2003 (first entry)
XX      DT      Saccharopolyspora butenyl-spinosyn biosynthetic gene cluster 2.
XX      DB      Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrolide;
XX      KW      metabolite; spinosyn; gene; ds.
```

```
XX OS Saccharopolyspora sp.
XX Key location/Qualifiers
FH CDS complement(114. .938)
FT /tag= a
FT /product= "bushF"
FT /note= "No start codon given"
FT 1389. .2561
FT /tag= b
FT /product= "bushG"
FT 2601. .3353
FT /tag= c
FT /product= "bushH"
FT complement(3359. .4546)
FT /tag= d
FT /product= "bushI"
FT 4684. .6303
FT /tag= e
FT /product= "bushJ"
FT 6317. .7510
FT /tag= f
FT /product= "bushK"
FT 7553. .8406
FT /tag= g
FT /product= "bushL"
FT /note= "No start codon given"
FT 8640. .9572
FT /tag= h
FT /product= "bushM"
FT /note= "No start codon given"
FT complement(9668. .10666)
FT /tag= i
FT /product= "bushN"
FT complement(10675. .12135)
FT /tag= j
FT /product= "bushO"
FT complement(12864. .14177)
FT /tag= k
FT /product= "bushP"
FT 14627. .15970
FT /tag= l
FT /product= "bushQ"
FT /note= "No start codon given"
FT 16008. .17144
FT /tag= m
FT /product= "bushR"
FT /note= "No start codon given"
FT 17168. .17917
FT /tag= n
FT /product= "bushS"
FT complement(18520. .19932)
FT /tag= o
FT /product= "ORF LI"
FT complement(19978. .20488)
FT /tag= p
FT /product= "ORF LIi"
FT /note= "No start codon given"
FT complement(20536. .21033)
FT /tag= q
FT /product= "ORF LIii"
FT /note= "No start codon given"
FT 21179. .21925
FT /tag= r
FT /product= "ORF LIiv"
FT complement(22671. .23453)
FT /tag= s
FT /product= "ORF LVI"
FT complement(23687. .24886)
FT /tag= t
FT /product= "ORF LVII"
FT complement(26177. .26923)
FT /tag= u
```

```
FT /product= "ORF LVIII"
FT /note= "No start codon given"
FT 27646. .28476
FT /tag= v
FT /product= "ORF LIX"
FT /note= "No start codon given"
FT
FT WO200279477-A2.
XX
XX 10-OCT-2002.
XX
XX 28-MAR-2002; 2002WO-US009968.
XX
XX 30-MAR-2001; 2001US-0280175P.
XX
XX (IDMC ) DOW AGROSCIENCES LLC.
XX
XX Hahn DR, Jackson JD, Bullard BS, Gustafson GD, Waldron C;
XX Mitchell JC;
XX
XX WPI; 2003-058434/05.
XX
XX P-P8DB; ABP57683, ABP57684, ABP57685, ABP57686, ABP57687, ABP57688.
XX
XX ABP57689, ABP57690, ABP57691, ABP57692, ABP57693, ABP57694, ABP57695,
XX
XX ABP57696, ABP57697, ABP57698, ABP57699, ABP57700, ABP57701, ABP57702,
XX
XX ABP57703, ABP57704.
XX
XX New butenyl-spinosyn biosynthetic genes, useful for increasing the
XX production of butenyl-spinosyn insecticidal macrocides, or for changing
XX the metabolites or products produced by spinosyn-producing
XX microorganisms.
XX
XX Claim 2; Page 99-119; 218pp; English.
XX
XX
XX The invention relates to a novel DNA molecule comprising a DNA sequence
XX that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn
XX PKS (polyketide synthetase) domain, or a spinosyn PKS module. The butenyl-
XX spinosyn biosynthetic genes are useful for increasing the production of
XX butenyl-spinosyn insecticidal macrocides. The genes are also useful for
XX changing the metabolites or products produced by spinosyn-producing
XX microorganisms. The present sequence represents a DNA molecule encoding
XX butenyl-spinosyn biosynthetic enzymes
XX
XX Sequence 36538 BP; 6867 A; 12266 C; 11182 G; 6223 T; 0 U; 0 Other;
XX
XX Query Match 6.9%; Score 57; DB 10; Length 36538;
XX Best Local Similarity 100.0%; Pred. No. 1.9e-17;
XX Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 73 TCGGTCGGGGGGGGCCCTGCGCCATCCACACGCGCTACTGGAGAAAGACGGGGCGG 129
XX |||||||
XX DB 866 TCGGTCGGGGGGGGCCCTGCGCCATCCACACGCGCTACTGGAGAAAGACGGGGCGG 810
XX
XX RESULT 6
XX ABV75562/c
XX ID ABV75562 standard; DNA; 22 BP.
XX
XX ABV75562;
XX
XX 22-JAN-2003 (first entry)
XX
XX Saccharopolyspora spnF reverse PCR primer.
XX
XX Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrocides;
XX metabolite; spinosyn; PCR; primer; ss; spnF.
XX
XX Saccharopolyspora spinosa.
XX
XX WO200279477-A2.
XX
XX 10-OCT-2002.
XX
XX 28-MAR-2002; 2002WO-US009968.
XX
XX PF
```



```

XX 30-MAR-2001; 2001US-0280175P.
PR (DOWC ) DOW AGROSCIENCES LLC.
XX
XX Hahn DR, Jackson JD, Bullard BS, Gustafson GD, Waldron C;
PI Mitchell JC;
XX WPI; 2003-058434/05.
XX
XX New butenyl-spinosyn biosynthetic genes, useful for increasing the
PT production of butenyl-spinosyn insecticidal macrolides, or for changing
PT the metabolites or products produced by spinosyn-producing
PT microorganisms.
XX
XX Example 1; Page 218; 218pp; English.
XX
XX The invention relates to a novel DNA molecule comprising a DNA sequence
CC that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn
CC PKS (polyketide synthetase) domain, or a spinosyn PKS module. The butenyl
CC -spinosyn biosynthetic genes are useful for increasing the production of
CC butenyl-spinosyn insecticidal macrolides. The genes are also useful for
CC changing the metabolites or products produced by spinosyn-producing
CC microorganisms. The present sequence represents a PCR primer used to
CC amplify the spinosyn biosynthetic gene spnF from a S. spinosa
CC biosynthetic gene cluster
XX
XX Sequence 22 BP; 3 A; 6 C; 5 G; 8 T; 0 U; 0 Other;
SQ
XX
XX Query Match 2.7%; Score 22; DB 10; Length 22;
XX Best Local Similarity 100.0%; Pred. No. 3.5;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 769 GAGAAATATGCCACGACATCG 790
DB 22 GAGAAATATGCCACGACATCG 1

```

# RESULT 7 AAQ70316/c

ID AAQ70316 standard; DNA; 457 BP.

XX AAQ70316;

DT 10-APR-1995 (first entry)

DE Oxidoreducing avermectin DNA from S. avermectilis FERM P-13155.

XX Modified; oxidoreducing; avermectin; S. avermectilis; wild type; disabled;  
KW de.

XX Streptomyces avermectilis.

OS JP06189774-A.

PN 12-JUL-1994.

PD 24-DEC-1992; 92JP-00344605.

PF 24-DEC-1992; 92JP-00344605.

PR (KITA ) KITASATO KENKYUSHO SH.

XX WPI; 1994-259554/32.

PT A new modified DNA fragment encoding an avermectin deriv. - useful for  
PT the prodn. of the deriv.

PS Claim 5; Page 2; 17pp; Japanese.

XX The sequences in AAQ80524 and AAQ70315-16 represent modified  
CC oxidoreducing avermectin genes from S. avermectilis strains deposited as  
ATCC 31271, FERM P-13154 and FERM P-13155. These genes were introduced

CC into S. avermectilis strains in which the wild type avermectin gene has  
CC been disabled such that the modified avermectins are produced in large  
CC amounts

SQ Sequence 457 BP; 83 A; 145 C; 146 G; 83 T; 0 U; 0 Other;

XX Query Match 2.7%; Score 22; DB 2; Length 457;

XX Best Local Similarity 100.0%; Pred. No. 2.8;

XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 143 AGCGCGCGACCGGCTCACCGA 164

DB 94 AGCGCGCGACCGGCTCACCGA 73

# RESULT 8 AAQ70315/c

ID AAQ70315 standard; DNA; 457 BP.

XX AAQ70315;

DT 10-APR-1995 (first entry)

DE Oxidoreducing avermectin DNA from S. avermectilis FERM P-13154.

XX Modified; oxidoreducing; avermectin; S. avermectilis; wild type; disabled;  
KW de.

OS Streptomyces avermectilis.

PN JP06189774-A.

PD 12-JUL-1994.

PF 24-DEC-1992; 92JP-00344605.

PR 24-DEC-1992; 92JP-00344605.

XX (KITA ) KITASATO KENKYUSHO SH.

XX WPI; 1994-259554/32.

PT A new modified DNA fragment encoding an avermectin deriv. - useful for  
PT the prodn. of the deriv.

PS Claim 4; Page 2; 17pp; Japanese.

XX The sequences in AAQ80524 and AAQ70315-16 represent modified  
CC oxidoreducing avermectin genes from S. avermectilis strains deposited as  
ATCC 31271, FERM P-13154 and FERM P-13155. These genes were introduced  
CC into S. avermectilis strains in which the wild type avermectin gene has  
CC been disabled such that the modified avermectins are produced in large  
CC amounts

SQ Sequence 457 BP; 83 A; 145 C; 145 G; 84 T; 0 U; 0 Other;

XX Query Match 2.7%; Score 22; DB 2; Length 457;

XX Best Local Similarity 100.0%; Pred. No. 2.8;

XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 143 AGCGCGCGACCGGCTCACCGA 164

DB 94 AGCGCGCGACCGGCTCACCGA 73

# RESULT 9 AAQ80524/c

ID AAQ80524 standard; DNA; 1410 BP.

XX AAQ80524;

DT 10-APR-1995 (first entry)

```
DE Oxidoreducing avermectin DNA from S. avermectilis ATCC31271.
XX Modified; oxidoreducing; avermectin; S. avermectilis; wild type; disabled;
KM ds.
XX Streptomyces avermectilis.
OS Key Location/Qualifiers
XX Key misc_feature 627..1075
XX FT /tag= a
XX FT /note= "Avermectin coding sequence"
XX FT
XX JP06189774-A.
XX 12-JUL-1994.
XX 24-DEC-1992; 92JP-00344605.
XX 24-DEC-1992; 92JP-00344605.
XX (KITA ) KITASATO KENKYUSHO SH.
XX MPI; 1994-259554/32.
XX A new modified DNA fragment encoding an avermectin deriv. - useful for
XX the prodn. of the deriv.
XX Claim 2; Page 2; 17pp; Japanese.
XX
XX The sequences in AA080524 and AA070315-16 represent modified
XX oxidoreducing avermectin genes from S. avermectilis strains deposited as
XX ATCC 31271, FERM P-13154 and FERM P-13155. These genes were introduced
XX into S. avermectilis strains in which the wild type avermectin gene has
XX been disabled such that the modified avermectins are produced in large
XX amounts
XX
XX Sequence 1410 BP; 232 A; 488 C; 463 G; 227 T; 0 U; 0 Other;
SQ
Query Match 2.7%; Score 22; DB 2; Length 1410;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 143 AGCGCGCGACCGGCTCACC GA 164
DB 720 AGCGCGCGACCGGCTCACC GA 639
RESULT 10
AA258381/c
ID AA258381 standard; DNA; 12381 BP.
XX
XX AA258381;
AC
XX
XX 23-MAY-2000 (first entry)
DT
XX Streptomyces avermectilis avermectin polyketide synthase modules 1+2.
DE
XX Polyketide synthase; avermectin; insecticide; ss.
XX
XX Streptomyces avermectilis.
OS
XX WO200001827-A2.
XX
XX 13-JAN-2000.
XX
XX 06-JUL-1999; 99WO-GB002158.
XX
XX 06-JUL-1998; 98GB-00014622.
XX
XX (BIOT-) BIOTICA TECHNOLOGY LTD.
XX (PFIZ ) PFIZER INC.
XX
XX Kellenberger JL, Leadley PF, Staunton J, Stutzman-Engwall KJ;
```

```
PI Mcarthur HAI;
XX
XX WPI; 2000-182117/16.
XX
XX Mutated Type I polyketide synthase containing a polylinker site in an
XX extension module for replacement of a reductive loop sequence, for
XX producing polyketides, e.g. BI avermectin.
XX
XX Disclosure; Fig 7a-f; 75pp; English.
XX
XX The present sequence is that of DNA encoding the first 2 modules of the
XX avermectin polyketide synthase (PKS) of Streptomyces avermectilis. The
XX invention relates to nucleic acids encoding a Type I PKS such as
XX avermectin in which a polylinker with multiple restriction sites replaces
XX or 1 more PKS genes encoding enzymes associated with reduction. Novel PKS
XX are provided in which in which the reductive loop in a selected module of
XX the Type I PKS is replaced with the equivalent segment from the same or
XX different PKS gene cluster or by a mutated or synthetic segment. Vectors
XX and host cells, and methods for producing novel polyketides by culturing
XX host cells are claimed. The polyketides obtained are useful as
XX antibiotics and insecticides. Fermentation products containing C22-C23
XX dihydroavermectin, ivermectin and BI avermectins are claimed
XX
XX Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 U; 0 Other;
SQ
Query Match 2.7%; Score 22; DB 3; Length 12381;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 143 AGCGCGCGACCGGCTCACC GA 164
DB 333 AGCGCGCGACCGGCTCACC GA 312
RESULT 11
ABV75561
ID ABV75561 standard; DNA; 20 BP.
XX
XX ABV75561;
AC
XX
XX 22-JAN-2003 (first entry)
DT
XX Saccharopolyspora spnF forward PCR primer.
XX
XX Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrolide;
XX metabolite; spinosyn; PCR; primer; ss; spnF.
XX
XX Saccharopolyspora spinosa.
XX
XX WO200279477-A2.
XX
XX 10-OCT-2002.
XX
XX 28-MAR-2002; 2002WO-US009968.
XX
XX 30-MAR-2001; 2001US-0280175P.
XX
XX (DOWC ) DOW AGROSCIENCES LLC.
XX
XX Hahn DR, Jackson JD, Bullard BS, Gustafson GD, Waldron C;
XX Mitchell JC;
XX
XX WPI; 2003-058434/05.
XX
XX New butenyl-spinosyn biosynthetic genes, useful for increasing the
XX production of butenyl-spinosyn insecticidal macrolides, or for changing
XX the metabolites or products produced by spinosyn-producing
XX microorganisms.
XX
XX Example 1; Page 217; 218pp; English.
XX
XX The invention relates to a novel DNA molecule comprising a DNA sequence
XX that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn
```

CC PKS (polyketide synthetase) domain, or a spinosyn PKS module. The butenyl  
CC -spinosyn biosynthetic genes are useful for increasing the production of  
CC butenyl-spinosyn insecticidal macrocides. The genes are also useful for  
CC changing the metabolites or products produced by spinosyn-producing  
CC microorganisms. The present sequence represents a PCR primer used to  
CC amplify the spinosyn biosynthetic gene spnf from a *S. spinosa*  
CC biosynthetic gene cluster  
XX  
SQ Sequence 20 BP; 6 A; 7 C; 5 G; 2 T; 0 U; 0 Other;  
Query Match 2.4%; Score 20; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 254 GCGACACGCGATCCGATC 273  
DB 1 GCGACACGCGATCCGATC 20  
RESULT 12  
ADP62503  
ID ADP62503 standard; cDNA; 237 BP.  
XX  
AC ADP62503;  
XX  
DT 09-SEP-2004 (first entry)  
XX  
DE Maize carbon assimilation pathway enzyme cDNA #2643.  
XX  
KM Carbon assimilation pathway enzyme; gene: ss; maize; corn;  
KM ribulose-bisphosphate carboxylase; phosphoglycerate kinase;  
KM glyceraldehyde 3-phosphate dehydrogenase; triose phosphate isomerase;  
KM aldolase; fructose-1,6-bisphosphate; transketolase;  
KM sedohepulose-1,7-bisphosphate; D-ribulose-5-phosphate kinase;  
KM ribose-5-phosphate isomerase; ribose-5-phosphate kinase;  
KM phosphoenolpyruvate carboxylase; NADP-dependent malate dehydrogenase;  
KM aspartate aminotransferase; alanine aminotransferase;  
KM NADP-dependent malic enzyme; NAD-dependent malic enzyme;  
KM PEP carboxykinase; pyruvate; phosphate dikinase; pyrophosphatase;  
KM plant metabolic pathway; plant breeding.  
XX  
OS Zea mays.  
XX  
PN US2004116682-A1.  
XX  
PD 17-JUN-2004.  
XX  
PF 16-NOV-2001; 2001US-00987899.  
XX  
PR 06-MAR-1998; 98US-0076912P.  
PR 04-MAR-1999; 99US-00262979.  
XX  
PA (CHEI/) CHEIKH N.  
PA (MILL/) MILLER P W.  
PA (OCON/) O'CONNELL K M.  
PA (LIU/) LIU J.  
PI Cheikh N, Miller PW, O'Connell KM, Liu J;  
PI WPI; 2004-498291/47.  
DR  
XX  
PT New substantially purified nucleic acid molecule encoding a maize or  
PT soybean carbon assimilation pathway enzyme, useful for gene mapping,  
PT identification and analysis, plant breeding, and preparation of  
PT constructs.  
XX  
PS Claim 2; SEQ ID NO 5572; 196pp; English.  
XX  
CC The invention relates to a substantially purified nucleic acid molecule  
CC that encodes a maize or soybean carbon assimilation pathway enzyme or its  
CC fragment. The maize or soybean carbon assimilation pathway enzyme or its  
CC fragment is selected from ribulose-bisphosphate carboxylase, triose  
CC phosphoglycerate kinase, glyceraldehyde 3-phosphate dehydrogenase, triose

CC phosphate isomerase, aldolase, fructose-1,6-bisphosphatase,  
CC transketolase, sedohepulose-1,7-bisphosphatase, D-ribulose-5-phosphate-3  
CC epimerase, ribose-5-phosphate isomerase, ribose-5-phosphate kinase,  
CC phosphoenolpyruvate carboxylase, NADP-dependent malate dehydrogenase,  
CC aspartate aminotransferase, alanine aminotransferase, NADP-dependent  
CC malic enzyme, NAD-dependent malic enzyme, PEP carboxykinase, pyruvate,  
CC phosphate dikinase and pyrophosphatase. The invention also relates to a  
CC substantially purified antibody or its fragment which is capable of  
CC specifically binding to a specific maize or soybean carbon assimilation  
CC pathway enzyme or its fragment, a transformed plant having a nucleic acid  
CC molecule comprising an exogenous promoter region which functions in a  
CC plant cell to cause the production of an mRNA molecule, and a method of  
CC determining a level or pattern in a plant cell of a carbon assimilation  
CC pathway enzyme in a plant metabolic pathway. The methods and compositions  
CC of the invention are useful for gene mapping, gene identification and  
CC analysis, plant breeding and preparation of constructs for use in plant  
CC gene expression and transgenic plants. This sequence represents cDNA  
CC encoding a maize carbon assimilation pathway enzyme of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
XX  
SQ Sequence 237 BP; 28 A; 101 C; 71 G; 37 T; 0 U; 0 Other;  
Query Match 2.4%; Score 20; DB 12; Length 237;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 492 CATCTCGGCGTACCAGG 511  
DB 192 CATCTCGGCGTACCAGG 211  
RESULT 13  
ADP62374  
ID ADP62374 standard; cDNA; 263 BP.  
XX  
AC ADP62374;  
XX  
DT 09-SEP-2004 (first entry)  
XX  
DE Maize carbon assimilation pathway enzyme cDNA #2514.  
XX  
KM Carbon assimilation pathway enzyme; gene: ss; maize; corn;  
KM ribulose-bisphosphate carboxylase; phosphoglycerate kinase;  
KM glyceraldehyde 3-phosphate dehydrogenase; triose phosphate isomerase;  
KM aldolase; fructose-1,6-bisphosphate; transketolase;  
KM sedohepulose-1,7-bisphosphate; D-ribulose-5-phosphate kinase;  
KM ribose-5-phosphate isomerase; ribose-5-phosphate kinase;  
KM phosphoenolpyruvate carboxylase; NADP-dependent malate dehydrogenase;  
KM aspartate aminotransferase; alanine aminotransferase;  
KM NADP-dependent malic enzyme; NAD-dependent malic enzyme;  
KM PEP carboxykinase; pyruvate; phosphate dikinase; pyrophosphatase;  
KM plant metabolic pathway; plant breeding.  
XX  
OS Zea mays.  
XX  
PN US2004116682-A1.  
XX  
PD 17-JUN-2004.  
XX  
PF 16-NOV-2001; 2001US-00987899.  
XX  
PR 06-MAR-1998; 98US-0076912P.  
PR 04-MAR-1999; 99US-00262979.  
XX  
PA (CHEI/) CHEIKH N.  
PA (MILL/) MILLER P W.  
PA (OCON/) O'CONNELL K M.  
PA (LIU/) LIU J.  
PI Cheikh N, Miller PW, O'Connell KM, Liu J;  
PI

DR WPI; 2004-498291/47.  
 XX  
 XX New substantially purified nucleic acid molecule encoding a maize or  
 PT soybean carbon assimilation pathway enzyme, useful for gene mapping, gene  
 PT identification and analysis, plant breeding, and preparation of  
 PT constructs.  
 PS  
 PS Claim 2; SEQ ID NO 5443; 196pp; English.  
 XX  
 XX The invention relates to a substantially purified nucleic acid molecule  
 CC that encodes a maize or soybean carbon assimilation pathway enzyme or its  
 CC fragment. The maize or soybean carbon assimilation pathway enzyme or its  
 CC fragment is selected from ribulose-bisphosphate carboxylase,  
 CC phosphoglycerate kinase, glyceraldehyde 3-phosphate dehydrogenase, triose  
 CC phosphate isomerase, aldolase, fructose-1,6-bisphosphatase, triose  
 CC transketolase, sedoheptulose-1,7-bisphosphatase, D-ribulose-5-phosphate-3  
 CC epimerase, ribose-5-phosphate isomerase, ribose-5-phosphate kinase,  
 CC phosphoenolpyruvate carboxylase, NADP-dependent malate dehydrogenase,  
 CC aspartate aminotransferase, alanine aminotransferase, NADP-dependent  
 CC malic enzyme, NAD-dependent malic enzyme, PEP carboxykinase, pyruvate,  
 CC phosphate dikinase and pyrophosphatase. The invention also relates to a  
 CC substantially purified antibody or its fragment which is capable of  
 CC specifically binding to a specific maize or soybean carbon assimilation  
 CC pathway enzyme or its fragment, a transformed plant having a nucleic acid  
 CC molecule comprising an exogenous promoter region which functions in a  
 CC plant cell to cause the production of an mRNA molecule, and a method of  
 CC determining a level or pattern in a plant cell of a carbon assimilation  
 CC pathway enzyme in a plant metabolic pathway. The methods and compositions  
 CC of the invention are useful for gene mapping, gene identification and  
 CC analysis, plant breeding and preparation of constructs for use in plant  
 CC gene expression and transgenic plants. This sequence represents cDNA  
 CC encoding a maize carbon assimilation pathway enzyme of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from USPTO at  
 CC seqdata.uspto.gov/sequence.html.  
 XX  
 XX Sequence 263 BP; 33 A; 105 C; 82 G; 43 T; 0 U; 0 Other;  
 SQ  
 Query Match 2.4%; Score 20; DB 12; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 492 CATCTCGGCGTCAACCGAGG 511  
 Db |||||  
 162 CATCTCGGCGTCAACCGAGG 181  
 RESULT 14  
 ADP62375 standard; cDNA; 290 BP.  
 ID ADP62375;  
 AC ADP62375;  
 XX  
 DT 09-SEP-2004 (first entry)  
 XX  
 DE Maize carbon assimilation pathway enzyme cDNA #2515.  
 XX  
 XX Carbon assimilation pathway enzyme; gene; ss; maize; corn;  
 KM ribulose-bisphosphate carboxylase; phosphoglycerate kinase;  
 KM glyceraldehyde 3-phosphate dehydrogenase; triose phosphate isomerase;  
 KM aldolase; fructose-1,6-bisphosphatase; transketolase;  
 KM sedoheptulose-1,7-bisphosphatase; D-ribulose-5-phosphate-3-epimerase;  
 KM ribose-5-phosphate isomerase; ribose-5-phosphate kinase;  
 KM phosphoenolpyruvate carboxylase; NADP-dependent malate dehydrogenase;  
 KM aspartate aminotransferase; alanine aminotransferase;  
 KM NADP-dependent malic enzyme; NAD-dependent malic enzyme;  
 KM PEP carboxykinase; pyruvate; phosphate dikinase; pyrophosphatase;  
 KM plant metabolic pathway; plant breeding.  
 XX  
 XX Zea mays.  
 XX US2004116682-A1.  
 XX

PD 17-JUN-2004.  
 XX  
 XX 16-NOV-2001; 2001US-00987899.  
 PF  
 XX 06-MAR-1998; 98US-0076912P.  
 PR  
 PR 04-MAR-1999; 99US-00262979.  
 XX  
 XX (CHEI/) CHEIKH N.  
 PA (MILL/) MILLER P W.  
 PA (OCON/) O'CONNELL K M.  
 PA (LIU/) LIU J.  
 XX  
 PI Cheikh N, Miller PW, O'Connell KM, Liu J,  
 WPI; 2004-498291/47.  
 DR  
 XX  
 XX New substantially purified nucleic acid molecule encoding a maize or  
 PT soybean carbon assimilation pathway enzyme, useful for gene mapping, gene  
 PT identification and analysis, plant breeding, and preparation of  
 PT constructs.  
 PS  
 PS Claim 2; SEQ ID NO 5444; 196pp; English.  
 XX  
 XX The invention relates to a substantially purified nucleic acid molecule  
 CC that encodes a maize or soybean carbon assimilation pathway enzyme or its  
 CC fragment. The maize or soybean carbon assimilation pathway enzyme or its  
 CC fragment is selected from ribulose-bisphosphate carboxylase,  
 CC phosphoglycerate kinase, glyceraldehyde 3-phosphate dehydrogenase, triose  
 CC phosphate isomerase, aldolase, fructose-1,6-bisphosphatase, triose  
 CC transketolase, sedoheptulose-1,7-bisphosphatase, D-ribulose-5-phosphate-3  
 CC epimerase, ribose-5-phosphate isomerase, ribose-5-phosphate kinase,  
 CC phosphoenolpyruvate carboxylase, NADP-dependent malate dehydrogenase,  
 CC aspartate aminotransferase, alanine aminotransferase, NADP-dependent  
 CC malic enzyme, NAD-dependent malic enzyme, PEP carboxykinase, pyruvate,  
 CC phosphate dikinase and pyrophosphatase. The invention also relates to a  
 CC substantially purified antibody or its fragment which is capable of  
 CC specifically binding to a specific maize or soybean carbon assimilation  
 CC pathway enzyme or its fragment, a transformed plant having a nucleic acid  
 CC molecule comprising an exogenous promoter region which functions in a  
 CC plant cell to cause the production of an mRNA molecule, and a method of  
 CC determining a level or pattern in a plant cell of a carbon assimilation  
 CC pathway enzyme in a plant metabolic pathway. The methods and compositions  
 CC of the invention are useful for gene mapping, gene identification and  
 CC analysis, plant breeding and preparation of constructs for use in plant  
 CC gene expression and transgenic plants. This sequence represents cDNA  
 CC encoding a maize carbon assimilation pathway enzyme of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from USPTO at  
 CC seqdata.uspto.gov/sequence.html.  
 XX  
 XX Sequence 290 BP; 38 A; 123 C; 74 G; 55 T; 0 U; 0 Other;  
 SQ  
 Query Match 2.4%; Score 20; DB 12; Length 290;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 492 CATCTCGGCGTCAACCGAGG 511  
 Db |||||  
 240 CATCTCGGCGTCAACCGAGG 259  
 RESULT 15  
 ADP62373 standard; cDNA; 294 BP.  
 ID ADP62373;  
 AC ADP62373;  
 XX  
 DT 09-SEP-2004 (first entry)  
 XX  
 DE Maize carbon assimilation pathway enzyme cDNA #2513.  
 XX  
 XX Carbon assimilation pathway enzyme; gene; ss; maize; corn;  
 KM ribulose-bisphosphate carboxylase; phosphoglycerate kinase;  
 KM

Search completed: August 29, 2005, 17:33:22  
JOD time : 591 secs

KW glyceraldehyde 3-phosphate dehydrogenase; triose phosphate isomerase;  
KW aldolase; fructose-1,6-bisphosphate; transketolase;  
KW sedohepulo-1,7-bisphosphate; D-ribulose-5-phosphate-3-epimerase;  
KW ribose-5-phosphate isomerase; ribose-5-phosphate kinase;  
KW phosphoenolpyruvate carboxylase; NADP-dependent malate dehydrogenase;  
KW aspartate aminotransferase; alanine aminotransferase;  
KW NADP-dependent malic enzyme; NAD-dependent malic enzyme;  
KW PEP carboxykinase; pyruvate; phosphate dikinase; pyrophosphatase;  
KW plant metabolic pathway; plant breeding.  
OS Zea mays.  
XX  
XX  
PN US2004116682-A1.  
XX  
PD 17-JUN-2004.  
XX  
PF 16-NOV-2001; 2001US-00987899.  
XX  
PR 06-MAR-1998; 98US-0076912P.  
XX  
PR 04-MAR-1999; 99US-00262979.  
XX  
PA (CHEI/) CHEIKH N.  
PA (MILL/) MILLER P W.  
PA (OCON/) O'CONNELL K M.  
PA (LIU/) LIU J.  
XX  
PI Cheikh N, Miller PW, O'Connell KM, Liu J;  
XX  
XX WPI; 2004-498291/47.  
XX  
PT New substantially purified nucleic acid molecule encoding a maize or  
PT soybean carbon assimilation pathway enzyme, useful for gene mapping, gene  
PT identification and analysis, plant breeding, and preparation of  
PT constructs.  
XX  
PS Claim 2; SEQ ID NO 5442; 196bp; English.  
XX  
CC The invention relates to a substantially purified nucleic acid molecule  
CC that encodes a maize or soybean carbon assimilation pathway enzyme or its  
CC fragment. The maize or soybean carbon assimilation pathway enzyme or its  
CC fragment is selected from ribulose-bisphosphate carboxylase,  
CC phosphoglycerate kinase, glyceraldehyde 3-phosphate dehydrogenase, triose  
CC phosphate isomerase, aldolase, fructose-1,6-bisphosphatase, transketolase,  
CC transketolase, sedohepulo-1,7-bisphosphatase, D-ribulose-5-phosphate-3-  
CC -epimerase, ribose-5-phosphate isomerase, ribose-5-phosphate kinase,  
CC phosphoenolpyruvate carboxylase, NADP-dependent malate dehydrogenase,  
CC aspartate aminotransferase, alanine aminotransferase, NADP-dependent  
CC malic enzyme, NAD-dependent malic enzyme, PEP carboxykinase, pyruvate,  
CC phosphate dikinase and pyrophosphatase. The invention also relates to a  
CC substantially purified antibody or its fragment which is capable of  
CC specifically binding to a specific maize or soybean carbon assimilation  
CC pathway enzyme or its fragment, a transformed plant having a nucleic acid  
CC molecule comprising an exogenous promoter region which functions in a  
CC plant cell to cause the production of an mRNA molecule, and a method of  
CC determining a level or pattern in a plant cell of a carbon assimilation  
CC pathway enzyme in a plant metabolic pathway. The methods and compositions  
CC of the invention are useful for gene mapping, gene identification and  
CC analysis, plant breeding and preparation of constructs for use in plant  
CC gene expression and transgenic plants. This sequence represents cDNA  
CC encoding a maize carbon assimilation pathway enzyme of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html.  
XX  
XX  
SQ Sequence 294 BP; 36 A; 127 C; 77 G; 54 T; 0 U; 0 Other;

## Query Match

2.4%; Score 20; DB 12; Length 294;

Best Local Similarity 100.0%; Fred. No. 27;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 492 CATCTCGGCTCACCGAGG 511

Db 249 CATCTCGGCTCACCGAGG 268

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OW nucleic - nucleic search, using SW model

Run on: August 29, 2005, 17:08:17 ; Search time 3416 Seconds  
(without alignments)  
9226.351 Million cell updates/sec

Title: US-10-069-353A-7

Sequence: 1 gtcgtgcgcagctgcgcacc.....cgcgaaagccgctgcgtcga 828

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 34239544 seqs, 19032134700 residues

Word size : 0

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gsa1:\*  
9: gb\_gsa2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	21	2.5	243	7	CK916588	CK916588 p3imgcf_0
2	21	2.5	436	7	CK917095	CK917095 p3imgcf_0
3	21	2.5	446	1	AU195670	AU195670 p3imgcf_0
4	21	2.5	495	7	CK916732	CK916732 p3imgcf_0
5	21	2.5	721	1	AU168919	AU168919 p3imgcf_0
6	21	2.5	892	9	CNS02017	AL208708 Tetraodon
7	20	2.4	122	2	AM799865	AM799865 PM1-UM005
8	20	2.4	213	5	BQ466968	BQ466968 HS02C10r
9	20	2.4	253	6	C23552	C23552 R1ce
10	20	2.4	309	1	AJ609797	AJ609797 AJ609797
11	20	2.4	341	1	AU070566	AU070566 AU070566
12	20	2.4	367	9	CC623049	CC623049 OGBR116TV
13	20	2.4	376	6	CD951480	CD951480 SAM_67 Ge
14	20	2.4	413	6	CD961336	CD961336 SDJ_109 G
15	20	2.4	453	6	CA501092	CA501092 WHE4029_B
16	20	2.4	453	1	AU181996	AU181996 AU181996
17	20	2.4	465	2	BE403165	BE403165 WHE0426_B
18	20	2.4	466	7	CO519448	CO519448 3530_1_12
19	20	2.4	475	5	BU982086	BU982086 HA5H18r
20	20	2.4	476	2	AM700130	AM700130 gB34h03.y
21	20	2.4	476	6	CB882644	CB882644 H102H08w
22	20	2.4	508	6	BG366814	BG366814 HVGME1000
23	20	2.4	517	5	BQ466703	BQ466703 HS01F02T
24	20	2.4	562	7	CO523560	CO523560 3530_1_15

25	20	2.4	567	6	CA137413	CA137413 SCAGRT203
26	20	2.4	580	6	CD998957	CD998957 QBF2h02.x
27	20	2.4	580	9	CR101195	CR101195 Reverse s
28	20	2.4	597	6	CA172082	CA172082 SCFSB107
29	20	2.4	597	6	CB883052	CB883052 R001A18w
30	20	2.4	613	6	CA067741	CA067741 SCQ5AD105
31	20	2.4	620	8	BZ716626	BZ716626 OGEAC41TM
32	20	2.4	621	5	BU102550	BU102550 SCCCAD100
33	20	2.4	621	6	CA064657	CA064657 SCCCAD100
34	20	2.4	621	6	CA014854	CA014854 HT12115r
35	20	2.4	629	8	BZ716620	BZ716620 OGEAC41TC
36	20	2.4	630	6	CD998956	CD998956 QBF2h02.p
37	20	2.4	630	6	CA014703	CA014703 HT12B20r
38	20	2.4	670	6	CD878187	CD878187 AZ04_102C
39	20	2.4	672	6	CA163453	CA163453 SCRLR2311
40	20	2.4	682	6	CA066408	CA066408 SCEQAD101
41	20	2.4	693	5	BQ466646	BQ466646 HS01C10T
42	20	2.4	693	5	CD872088	CD872088 AZ02_119M
43	20	2.4	695	6	CA185154	CA185154 SCBSST309
44	20	2.4	712	6	CD440782	CD440782 EL01N0560
45	20	2.4	758	6	CA130606	CA130606 SCCCRT100

#### ALIGNMENTS

RESULT 1  
CK916588  
LOCUS 243 bp mRNA linear EST 22-APR-2004  
DEFINITION p3imgcf\_003617 Normalized Magnaporthe grisea cDNA PRS423 library  
CK916588  
MAGNAPORTHE GRISEA cDNA clone p3imgcf\_003617, mRNA sequence.

ACCESSION  
CK916588  
VERSION  
CK916588.1 GI:45375293

KEYWORDS  
EST.  
MAGNAPORTHE GRISEA (anamorph: Pyricularia grisea)  
MAGNAPORTHE GRISEA  
Bukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

ORGANISM  
Sordariomycetes; Incertae sedis; Magnaporthaceae; Magnaporthe.  
1 (bases 1 to 243)

REFERENCE  
Dong, H., Peng, Y., Chen, B., Li, Y. and Li, D.  
large-scale identification of ESTs from Magnaporthe grisea by  
normalized cDNA library sequencing (2004b)

#### JOURNAL

Unpublished (2004)

#### COMMENT

Contact: Debao Li  
Bioinformatics and Gene Network Research Group  
Zhejiang University, China Agricultural University, Guangxi

Kaixuan Road 268#, Hangzhou, Zhejiang, 310029, china

Tel: 0086-571-86961525

Fax: 0086-571-86971183

Email: webmaster@estarray.org, URL: http://www.estarray.org

Seq primer: M13 reverse primer.

#### FEATURES

Location/Qualifiers

#### source

1..243  
/organism="Magnaporthe grisea"  
/mol\_type="mRNA"  
/db\_xref="taxon:148305"  
/clone="p3imgcf\_003617"  
/issue\_type="Mycelium, conidium, germinating conidium,  
swelling appressorium, mature appressorium, penetration  
peg"  
/dev\_stage="Mycelium, conidium, germinating conidium,  
swelling appressorium, mature appressorium, penetration  
peg"  
/clone\_lib="Normalized Magnaporthe grisea cDNA PRS423  
library"  
/note="Vector: PRS423"

#### ORIGIN

Query Match 2.5%; Score 21; DB 7; Length 243;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 CAACGAGAGCGGGCGGG 536  
 |||||  
 Db 77 CAACGAGAGCGGGCGGG 97

RESULT 2 436 bp mRNA linear EST 22-APR-2004  
 LOCUS CK917095  
 DEFINITION p3fmgcf\_004212 Normalized Magnaporthe grisea cDNA PRS423 library  
 Magnaporthe grisea clone p3fmgcf\_004212, mRNA sequence.

ACCESSION CK917095  
 VERSION CK917095.1 GI:45375800  
 KEYWORDS EST.  
 SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)  
 ORGANISM Magnaporthe grisea  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.  
 1 (bases 1 to 436)  
 REFERENCE Dong,H., Peng,Y., Chen,B., Li,Y. and Li,D.  
 Large-scale identification of ESTs from Magnaporthe grisea by  
 normalized cDNA library sequencing (2004b)  
 JOURNAL Unpublished (2004)  
 COMMENT Contact: Debao Li  
 Bioinformatics and Gene Network Research Group  
 Zhejiang University, China Agricultural University, Guangxi  
 University  
 Kaixuan Road 268#, Hangzhou, Zhejiang, 310029, china  
 Tel: 0086-571-86961525  
 Fax: 0086-571-86971183  
 Email: webmaster@estarray.org, URL: http://www.estarray.org  
 Seq primer: M13 reverse primer.  
 Location/Qualifiers  
 1..436  
 /organism="Magnaporthe grisea"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:148305"  
 /clone="p3fmgcf\_004212"  
 /tissue\_type="Mycelium, conidium, germinating conidium,  
 swelling appressorium, mature appressorium, penetration  
 peg"  
 /dev\_stage="Mycelium, conidium, germinating conidium,  
 swelling appressorium, mature appressorium, penetration  
 peg"  
 /clone\_lib="Normalized Magnaporthe grisea cDNA PRS423  
 library"  
 /note="Vector: pRS423"

ORIGIN

Query Match 2.5%; Score 21; DB 7; Length 436;  
 Best Local Similarity 100.0%; Pred. No. 9.9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 CAACGAGAGCGGGCGGG 536  
 |||||  
 Db 269 CAACGAGAGCGGGCGGG 289

RESULT 3 446 bp mRNA linear EST 14-OCT-2003  
 LOCUS AU195670/c  
 DEFINITION AU195670 Porphyra yezoensis TU-1 sporophytes Porphyra yezoensis  
 cDNA clone PFL078F08\_r\_5', mRNA sequence.

ACCESSION AU195670  
 VERSION AU195670.1 GI:31937533  
 KEYWORDS EST.  
 SOURCE Porphyra yezoensis  
 ORGANISM Porphyra yezoensis  
 Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;  
 Porphyra.  
 1 (bases 1 to 446)  
 REFERENCE Asamizu,E., Nakajima,M., Kitade,Y., Saga,N., Nakamura,Y. and  
 Tabata,S.  
 COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS

OF PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG  
 FREQUENCY ANALYSIS  
 J. Phycol. 39 (5), 923-930 (2003)  
 JOURNAL Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yama 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
 source  
 1..446  
 /organism="Porphyra yezoensis"  
 /mol\_type="mRNA"  
 /strain="TU-1"  
 /db\_xref="taxon:2788"  
 /clone="PFL078F08\_r"  
 /dev\_stage="sporophytes"  
 /clone\_lib="Porphyra yezoensis TU-1 sporophytes"

ORIGIN

Query Match 2.5%; Score 21; DB 1; Length 446;  
 Best Local Similarity 100.0%; Pred. No. 9.9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 GGCGTCACCGAGTCGTCAA 519  
 |||||  
 Db 99 GGCGTCACCGAGTCGTCAA 79

RESULT 4 495 bp mRNA linear EST 22-APR-2004  
 LOCUS CK916732  
 DEFINITION p3fmgcf\_003787 Normalized Magnaporthe grisea cDNA PRS423 library  
 Magnaporthe grisea clone p3fmgcf\_003787, mRNA sequence.

ACCESSION CK916732  
 VERSION CK916732.1 GI:45375437  
 KEYWORDS EST.  
 SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)  
 ORGANISM Magnaporthe grisea  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.  
 1 (bases 1 to 495)  
 REFERENCE Dong,H., Peng,Y., Chen,B., Li,Y. and Li,D.  
 Large-scale identification of ESTs from Magnaporthe grisea by  
 normalized cDNA library sequencing (2004b)  
 JOURNAL Unpublished (2004)  
 COMMENT Contact: Debao Li  
 Bioinformatics and Gene Network Research Group  
 Zhejiang University, China Agricultural University, Guangxi  
 University  
 Kaixuan Road 268#, Hangzhou, Zhejiang, 310029, china  
 Tel: 0086-571-86961525  
 Fax: 0086-571-86971183  
 Email: webmaster@estarray.org, URL: http://www.estarray.org  
 Seq primer: M13 reverse primer.  
 Location/Qualifiers  
 1..495  
 /organism="Magnaporthe grisea"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:148305"  
 /clone="p3fmgcf\_003787"  
 /tissue\_type="Mycelium, conidium, germinating conidium,  
 swelling appressorium, mature appressorium, penetration  
 peg"  
 /dev\_stage="Mycelium, conidium, germinating conidium,  
 swelling appressorium, mature appressorium, penetration  
 peg"  
 /clone\_lib="Normalized Magnaporthe grisea cDNA PRS423  
 library"  
 /note="Vector: pRS423"

ORIGIN

Query Match 2.5%; Score 21; DB 7; Length 495;  
 Best Local Similarity 100.0%; Pred. No. 9.8;



Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 CAACGAGAGCGCGCGCGG 536  
 |||||  
 332 CAACGAGAGCGCGCGCGG 352

RESULT 5  
 AUI68919/c 721 bp mRNA linear EST 29-JAN-2001  
 LOCUS AUI68919 OI-br-ad cDNA Oryzias latipes cDNA clone br3805, mRNA  
 DEFINITION sequence.

ACCESSION AUI68919  
 VERSION AUI68919.1 GI:12590988  
 KEYWORDS EST.  
 SOURCE Oryzias latipes (Japanese medaka)  
 ORGANISM Oryzias latipes

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 TITLE Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 JOURNAL Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 COMMENT Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.  
 1 (bases 1 to 721)  
 Mita, K., Ishikawa, Y. and Yamauchi, M.  
 Establishment of cDNA database of medaka, Oryzias latipes  
 Unpublished (2001)  
 CONTACT: Malta K  
 Genome Research Group  
 National Institute of Agrobiological Sciences  
 Otsu 1-2, Tsukuba, Ibaraki 305-8634, Japan  
 Email: kmika@nias.affrc.go.jp  
 method: uni-directional sequence direction: sequenced from T3 primer  
 (5' -> 3').

FEATURES  
 source  
 1. 721  
 /organism="Oryzias latipes"  
 /mol\_type="mRNA"  
 /ecrln="HNT"  
 /db\_xref="taxon:8090"  
 /clone="br3805"  
 /sex="female/male mixed"  
 /tissue\_type="brain"  
 /dev\_stage="adult"  
 /clone\_1ib="OI-br-ad cDNA"

ORIGIN  
 Query Match 2.5%; Score 21; DB 1; Length 721;  
 Best Local Similarity 100.0%; Pred. No. 9.6;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 GGGCTTCTGGCAGCAGCGCG 148  
 |||||  
 583 GGGCTTCTGGCAGCAGCGCG 563

RESULT 6  
 CNGS02Q17 892 bp DNA linear GSS 01-SEP-2000  
 LOCUS CNGS02Q17  
 DEFINITION Tetradon nigroviridis genome survey sequence PUC-ORI end of clone  
 156N1 of library G from Tetradon nigroviridis, genomic survey  
 sequence.

ACCESSION AL208708.1 GI:7867527  
 VERSION AL208708.1  
 KEYWORDS GSS; genome survey sequence.  
 SOURCE Tetradon nigroviridis  
 ORGANISM Tetradon nigroviridis

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 TITLE Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 JOURNAL Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 COMMENT Tetraodontidae; Tetraodon.  
 1  
 Roest Crolius, H., Jailon, O., Dasilva, C., Bouneau, L., Fisher, C.,  
 Bernot, A., Fizames, C., Mincker, P., Brottier, P., Quetier, F.,  
 Saurin, W. and Weissenbach, J.

TITLE Estimate of human gene number provided by genome-wide analysis  
 using Tetradon nigroviridis DNA sequence  
 JOURNAL Nat. Genet. 25 (2), 235-238 (2000)  
 MEDLINE 20296633  
 PUBMED 10835645  
 REFERENCE 2

Roest Crolius, H., Jailon, O., Dasilva, C., Ozouf-Costaz, C.,  
 Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,  
 Saurin, W., Bernot, A. and Weissenbach, J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetradon nigroviridis  
 Genome Res. 10 (7), 939-949 (2000)

JOURNAL 20359837  
 MEDLINE 10899143  
 PUBMED 3 (bases 1 to 892)

REFERENCE  
 Genoscope.  
 Direct Submission  
 Submitted (12-Apr-2000) Genoscope - Centre National de Séquençage :  
 BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetradon nigroviridis  
 genome. For more information, please take a look at  
 http://www.genoscope.cns.fr/Tetradon.

FEATURES  
 source  
 1. 892  
 /organism="Tetradon nigroviridis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:99883"  
 /clone="156N10"  
 /clone\_1ib="G"  
 /note="TGenoscope sequence ID : C0AG156DG05SP1-end :  
 PUC-ORI"

ORIGIN  
 Query Match 2.5%; Score 21; DB 9; Length 892;  
 Best Local Similarity 100.0%; Pred. No. 9.5;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 543 GGTGTCCGGGAGCAGGTGCC 563  
 |||||  
 653 GGTGTCCGGGAGCAGGTGCC 673

RESULT 7  
 AW799865 122 bp mRNA linear EST 16-MAY-2000  
 LOCUS PML-UM0056-100300-003-f06 UM0056 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION AW799865  
 VERSION AW799865.1 GI:7851735  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 122)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
 Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V.,  
 O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.U.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL 20202663  
 MEDLINE 10737800  
 PUBMED

COMMENT  
 Contact: Simpson A.U.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922

Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the PAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL:  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=ct2=PM1-UM0056-100  
 300-003-f06&ct=2000-03-10&cl=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 7  
 High quality sequence stop: 71.  
 Location/Qualifiers

## FEATURES

source

```
1..122
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="UM0056"
/note="Organ: uterus; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
```

## ORIGIN

Query Match 2.4%; Score 20; DB 2; Length 122;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 604 TCCTGCGGTGACGCGGGTT 623  
 ||||||||||||  
 Db 99 TCCTGCGGTGACGCGGGTT 118

## RESULT 8

BQ466968

LOCUS BQ466968 213 bp mRNA linear EST 30-MAY-2002  
 DEFINITION HS02C10r HS Hordeum vulgare subsp. vulgare cDNA clone HS02C10

5-PRIME, mRNA sequence.  
 BQ466968

ACCESSION BQ466968.1 GI:21274750  
 EST.

## KEYWORDS

SOURCE

ORGANISM

Hordeum vulgare subsp. vulgare  
 Hordeum vulgare subsp. vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooideae; Triticeae; Hordeum.  
 1 (bases 1 to 213)

## REFERENCE

AUTHORS

Zhang, H., Potokina, E., Michalek, W., Weschke, W., Stein, N. and  
 Graner, A.  
 Barley ESTs from germinating seeds  
 Unpublished (2002)

## TITLE

JOURNAL

Contact: Stein Nils  
 Molecular Markers Group, Department Genbank  
 Institute of Plant Genetics and Crop Plant Research (IPK)  
 Corrensstr. 3, 06466, Gatersleben, Germany  
 Tel: 039482-5522  
 Fax: 039482-5595

## COMMENT

Email: stein@ipk-gatersleben.de  
 Insert Length: 213 Std Error: 0.00  
 Plate: 2 row: C column: 10  
 Seq primer: M3rev.  
 Location/Qualifiers

## FEATURES

source

```
1..213
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Darke"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HS02C10"
/tissue_type="embryo + scutellum"
/dev_stage="0-16 hours after imbibition"
/lab_host="XL10-Gold"
```

/clone\_lib="HS"  
 /note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/White selection for recombinants is not 100% reliable."

## ORIGIN

Query Match 2.4%; Score 20; DB 5; Length 213;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 492 CATCTCGGCGTCAACCGAG 511  
 ||||||||||||  
 Db 168 CATCTCGGCGTCAACCGAG 187

## RESULT 9

C23552

LOCUS C23552 253 bp mRNA linear EST 03-APR-2002  
 DEFINITION C23552 Rice callus Oryza sativa (japonica cultivar-group) cDNA

clone C60213\_1A, mRNA sequence.  
 C23552

ACCESSION C23552.1 GI:1944230  
 EST.

## KEYWORDS

SOURCE

ORGANISM

Oryza sativa (japonica cultivar-group)  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoidae; Oryzaceae; Oryza.  
 1 (bases 1 to 253)

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki  
 305-8602, Japan  
 Tel: 81-288-38-7441  
 Fax: 81-288-38-7468  
 Email: tsasaki@db.affrc.go.jp, URL: http://rsgp.dna.affrc.go.jp/  
 PROJECT = 'RGP'.  
 Location/Qualifiers

## FEATURES

source

```
1..253
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="C60213_1A"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site 1: SalI; Site 2: NotI; cDNA prepared from rice callus mRNAs by using oligo(dT) as a primer and ligating to the SalI-NotI site of pBluescript II SK+ phagemid."
```

## ORIGIN

Query Match 2.4%; Score 20; DB 6; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 492 CATCTCGGCGTCAACCGAG 511  
 ||||||||||||  
 Db 147 CATCTCGGCGTCAACCGAG 166

## RESULT 10

AJ609797

LOCUS AJ609797 309 bp mRNA linear EST 11-DEC-2003  
 DEFINITION AJ609797 Triticum turgidum subsp. durum etiolated seedling 20 day  
 Triticum turgidum subsp. durum cDNA clone 00262R, mRNA sequence.

ACCESSION AJ609797  
 VERSION AJ609797.1 GI:39726724  
 KEYWORDS EST.  
 SOURCE Triticum turgidum subsp. durum (durum wheat)  
 ORGANISM Triticum turgidum subsp. durum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 1 (bases 1 to 309)  
 Cifarelli, R.A., D'Onofrio, O., and Lauria, G.  
 Expressed Sequences Tags (ESTs) library from totipotent cDNA of  
 durum wheat  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Cifarelli RA  
 Biotechnology  
 Metapontum Agrobios  
 S.S. Jonica 106 Km 448.2, 75010 Metaponto (MT), Italy.  
 Location/Qualifiers  
 1..309  
 /organism="Triticum turgidum subsp. durum"  
 /mol\_type="mRNA"  
 /cultivar="Ofanto"  
 /sub\_species="durum"  
 /db\_xref="taxon:4567"  
 /clone="00262R"  
 /issue\_type="etiolated seedling"  
 /dev\_stage="20 day"  
 /clone\_lib="Triticum turgidum subsp. durum etiolated  
 seedling 20 day"

ORIGIN  
 Query Match 2.4%; Score 20; DB 1; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 492 CATCTCGGCGTCACCGAG 511  
 |||  
 200 CATCTCGGCGTCACCGAG 219

RESULT 11  
 LOCUS AU070566 341 bp mRNA linear EST 02-APR-2002  
 DEFINITION AU070566 Rice shoot Oryza sativa (japonica cultivar-group) cDNA  
 clone S4279\_8A, mRNA sequence.  
 ACCESSION AU070566  
 VERSION AU070566.1 GI:5005404  
 KEYWORDS EST.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 1 (bases 1 to 341)  
 Yamamoto, K. and Sasaki, T.  
 Rice cDNA from etiolated shoot  
 Unpublished (1997)  
 Contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki  
 305-8602, Japan  
 Tel: 81-298-38-7441  
 Fax: 81-298-38-7468  
 Email: tsasaki@affrc.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/.  
 location/Qualifiers  
 1..341  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="S4279\_8A"  
 /dev\_stage="Etiolated shoot (8 days old)"  
 /clone\_lib="Rice shoot"

FEATURES  
 source

ORIGIN /note="Etiolated shoot (8 days old)"  
 Query Match 2.4%; Score 20; DB 1; Length 341;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 492 CATCTCGGCGTCACCGAG 511  
 |||  
 107 CATCTCGGCGTCACCGAG 126

RESULT 12  
 LOCUS CC623049/c 367 bp DNA linear GSS 19-JUN-2003  
 DEFINITION OGRB1cTV ZM 0.7.1.5 KB Zea mays genomic clone ZMMBma0380D08,  
 genomic survey sequence.  
 ACCESSION CC623049  
 VERSION CC623049.1 GI:31989470  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoidae; Andropogoneae; Zea.  
 1 (bases 1 to 367)  
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Uterback, T.,  
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, D.A., Rohlfing, T.,  
 Citek, R.W., Nurnberg, A., Robbins, D., and Lakey, N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 Contact: Cathy Whitelaw  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TF  
 Classes: sheared ends.  
 Location/Qualifiers  
 1..367  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBma0380D08"  
 /clone\_lib="ZM 0.7.1.5 KB"  
 /note="Vector: pBCSK-; Site 1: HincII, 0.7-1.5 kb  
 methylation filtered genomic DNA library"

ORIGIN  
 Query Match 2.4%; Score 20; DB 9; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 CGTCATGCCATGTCCCTGC 382  
 |||  
 148 CGTCATGCCATGTCCCTGC 129

RESULT 13  
 LOCUS CD951480 376 bp mRNA linear EST 15-JUL-2003  
 DEFINITION SAW 67 GeneTag2 Zea mays cDNA, mRNA sequence.  
 ACCESSION CD951480  
 VERSION CD951480.1 GI:32799244  
 KEYWORDS EST.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoidae; Andropogoneae; Zea.  
 1 (bases 1 to 376)

REFERENCE

AUTHORS Genoplante.  
TITLE Genoplante, a major partnership french program in plant genomics  
JOURNAL Unpublished (2003)  
COMMENT Contact: Genoplante  
Genoplante  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french plant genome programme 'Genoplante' (<http://www.genoplante.com> and <http://genoplante-info.infobiogen.fr>).  
FEATURES  
source  
1..376  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="mixture"  
/db\_xref="taxon:4577"  
/clone\_lib="GenetAg2"  
ORIGIN  
Query Match 2.4%; Score 20; DB 6; Length 376;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 363 CGTCATGCCATGTCCTGC 382  
|||||  
Db 215 CGTCATGCCATGTCCTGC 234  
RESULT 14  
LOCUS CD961336 376 bp mRNA linear EST 15-JUL-2003  
DEFINITION SD\_109 GeneTag2 Zea mays CDNA, mRNA sequence.  
ACCESSION CD961336  
VERSION CD961336.1 GI:32809102  
KEYWORDS EST.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 376)  
REFERENCE  
AUTHORS Genoplante.  
TITLE Genoplante, a major partnership french program in plant genomics  
JOURNAL Unpublished (2003)  
COMMENT Contact: Genoplante  
Genoplante  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french plant genome programme 'Genoplante' (<http://www.genoplante.com> and <http://genoplante-info.infobiogen.fr>).  
FEATURES  
source  
Location/Qualifiers  
1..376  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="mixture"  
/db\_xref="taxon:4577"  
/clone\_lib="GenetAg2"  
ORIGIN  
Query Match 2.4%; Score 20; DB 6; Length 376;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 363 CGTCATGCCATGTCCTGC 382  
|||||  
Db 215 CGTCATGCCATGTCCTGC 234  
RESULT 15  
CAS01092

LOCUS CAS01092 413 bp mRNA linear EST 14-NOV-2002  
DEFINITION WHE4029\_B12\_C23T Wheat meiotic anther cDNA library Triticum aestivum CDNA clone WHE4029\_B12\_C23, mRNA sequence.  
ACCESSION CAS01092  
VERSION CAS01092.1 GI:24992052  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Triticum.  
1 (bases 1 to 413)  
REFERENCE  
AUTHORS Anderson,O.D., Chao,S., Crossman,C., Langridge,P., Iazo,G.R., Pham,V., Rausch,C.J., Sutton,T., Woo,J. and Wilson,C.  
TITLE The structure and function of the expressed portion of the wheat genomes - Meiotic anther cDNA library  
JOURNAL Unpublished (2002)  
COMMENT Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: oanderson@w.usda.gov  
Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20  
Seq primer: T7 primer.  
FEATURES  
source  
Location/Qualifiers  
1..413  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="WHE4029\_B12\_C23"  
/tissue\_type="Anther"  
/dex\_stage="Meiotic stages pre-meiosis-metaphase I"  
/lab\_host="E. coli DH10B"  
/clone\_lib="Wheat meiotic anther cDNA library"  
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; Plants were grown in a glasshouse. Anther meiotic stage was determined by removing anthers from individual primary florets. One anther was sacrificed for microscopic staging, and if determined to be between (and including) meiotic stages pre-meiosis and metaphase I, the remaining two anthers were collected and pooled for library construction. The tissue, total RNA, and poly(A) RNA were prepared, cDNA synthesised, and directionally ligated into pSPORT1 by Tim Sutton in the P Langridge lab at the Department of Plant Science, University of Adelaide, Waite Campus, Australia. Average insert size 1.5kb. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."  
ORIGIN  
Query Match 2.4%; Score 20; DB 6; Length 413;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 492 CATCTCGGCGTCACGAGG 511  
|||||  
Db 182 CATCTCGGCGTCACGAGG 201  
Search completed: August 29, 2005, 19:35:30  
Job time : 3428 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: August 29, 2005, 17:11:57 ; Search time 219 Seconds  
(without alignments)  
6186.471 Million cell updates/sec

Title: US-10-069-353A-7

Perfect score: 828  
Sequence: 1 ggcgtgcgcagtcgagcacc.....cgcggaagccggtcgcgtga 828

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1202784 seqs, 818138359 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/prodata/1/ina/PCUS\_COMB.seq: \*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	DB	ID	Description
1	828	100.0	80161	3	US-09-036-987A-1
2	828	100.0	80161	3	US-09-370-700-1
3	828	100.0	80161	4	US-09-603-207-1
4	20	2.4	3107	2	US-08-813-940-3
5	20	2.4	8051	2	US-08-576-626A-2
6	20	2.4	4403765	3	US-09-103-840A-2
7	20	2.4	4411529	3	US-09-103-840A-1
8	19	2.3	1330025	4	US-09-198-452A-1
9	19	2.3	1330230	4	US-09-438-185A-1
10	18	2.2	266	4	US-09-313-294A-3098
11	18	2.2	272	4	US-09-313-294A-2444
12	18	2.2	282	4	US-09-621-976-489
13	18	2.2	391	4	US-09-513-999C-10082
14	18	2.2	429	4	US-09-252-991A-1854
15	18	2.2	513	4	US-09-252-991A-1154
16	18	2.2	636	4	US-09-170-984-6
17	18	2.2	897	4	US-09-252-991A-9787
18	18	2.2	939	4	US-09-252-991A-10191
19	18	2.2	945	4	US-09-252-991A-2181
20	18	2.2	1194	4	US-09-252-991A-1785
21	18	2.2	1203	4	US-09-252-991A-2018
22	18	2.2	1386	4	US-09-602-787A-359
23	18	2.2	1536	4	US-09-180-109A-3
24	18	2.2	1818	4	US-09-252-991A-1706
25	18	2.2	2052	4	US-09-252-991A-10050
26	18	2.2	2568	4	US-09-489-039A-2980
27	18	2.2	3981	4	US-09-252-991A-1361

C 28	18	2.2	4179	4	US-09-252-991A-1244	Sequence 1244, Ap
C 29	17	2.1	204	4	US-09-252-991A-11819	Sequence 11819, A
C 30	17	2.1	256	4	US-09-270-767-7360	Sequence 7360, Ap
C 31	17	2.1	256	4	US-09-270-767-22642	Sequence 22642, A
C 32	17	2.1	428	4	US-09-513-999C-2346	Sequence 2346, Ap
C 33	17	2.1	482	4	US-09-621-976-3490	Sequence 3490, Ap
C 34	17	2.1	501	4	US-09-902-540-3908	Sequence 3908, Ap
C 35	17	2.1	531	4	US-09-252-991A-7115	Sequence 7115, Ap
C 36	17	2.1	537	4	US-09-252-991A-5582	Sequence 5582, Ap
37	17	2.1	573	4	US-09-902-540-7223	Sequence 7223, Ap
38	17	2.1	580	1	US-08-078-176-1	Sequence 1, Appl
39	17	2.1	684	4	US-09-252-991A-2861	Sequence 2861, Ap
C 40	17	2.1	717	4	US-09-902-540-7027	Sequence 7027, Ap
C 41	17	2.1	792	4	US-09-252-991A-7444	Sequence 7444, Ap
42	17	2.1	852	4	US-09-266-965-36	Sequence 36, Appl
43	17	2.1	888	4	US-09-252-991A-7041	Sequence 7041, Ap
44	17	2.1	927	4	US-09-252-991A-1851	Sequence 1851, Ap
45	17	2.1	960	4	US-09-252-991A-7256	Sequence 7256, Ap

#### ALIGNMENTS

RESULT 1  
US-09-036-987A-1  
; Sequence 1, Application US/09036987A  
; Patent No. 6143526  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H.  
; APPLICANT: Broughton, Mary C.  
; APPLICANT: Crawford, Kathryn P.  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Merlo, Donald J.  
; APPLICANT: Turner, Jan R.  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; TITLE OF INVENTION: Production  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dow Agrosciences LLC Patent Department  
; STREET: 9330 Zionville Road  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: USA  
; Zip: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Releasee #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/036,987A  
; FILING DATE: 09-MAR-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stuart, Donald R  
; REGISTRATION NUMBER: 28,479  
; REFERENCE/DOCKET NUMBER: 50, 608  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (317)337-4816  
; TELEFAX: (317)337-4847  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 80161 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-09-036-987A-1  
Query Match 100.0%; Score 828; DB 3; Length 80161;  
Best Local Similarity 100.0%; Pred. No. 0;

```
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGTGGCCAGGTGGCCGCAACCATGACGAGGTGGGGAATGATGACCTGGTCAG 60
Db 20168 GTTTCACAGGTGGCCGCAACCATGACGAGGTGGGGAATGATGACCTGGTCAG 20227
QY 61 CCGTGTCTGAATCTGGTGGGCGGCGCCCTGGCCATCCACGAGCTTACTGGAGAAC 120
Db 20228 CCGTGTCTGAATCTGGTGGGCGGCGCCCTGGCCATCCACGAGCTTACTGGAGAAC 20287
QY 121 GACGGGCGGGCTTCTGTCGACGAGGCGCGCCGACCGGCTACCGACCTTGTCCGAAACG 180
Db 20288 GACGGGCGGGCTTCTGTCGACGAGGCGCGCCGACCGGCTACCGACCTTGTCCGAAACG 20347
QY 181 ACCGTGCTGATGAGCGGCGCTGACTGCTGATGATGGGGTGGGCTACCGACCAACGAGCG 240
Db 20348 ACCGTGCTGATGAGCGGCGCTGACTGCTGATGATGGGGTGGGCTACCGACCAACGAGCG 20407
QY 241 CTGGCGGTGCGCGCGACCAACCGGATCCAGATCCAGGATCAACCGTCAAGCGAGTGCAA 300
Db 20408 CTGGCGGTGCGCGCGACCAACCGGATCCAGATCCAGGATCAACCGTCAAGCGAGTGCAA 20467
QY 301 GTGGCCATGCGCGCTGATGTCGCGGACCGGAACTAAGCAACCGGCTGAGACTTCTG 360
Db 20468 GTGGCCATGCGCGCTGATGTCGCGGACCGGAACTAAGCAACCGGCTGAGACTTCTG 20527
QY 361 TCGGTGATGATGATGCTCCCTGCGATCCGACCAATGCTTTGACGCGCGCTGGGCGCATG 420
Db 20528 TCGGTGATGATGATGCTCCCTGCGATCCGACCAATGCTTTGACGCGCGCTGGGCGCATG 20587
QY 421 CAGTGCCTGTGGAGATGTCGGAACCGGACCGGTCCATCCGGAATCTCTTGAAGTACTC 480
Db 20588 CAGTGCCTGTGGAGATGTCGGAACCGGACCGGTCCATCCGGAATCTCTTGAAGTACTC 20647
QY 481 AAACCCGGGTGGCATCTCTCGCGGTGACCCGAGTCTGTAACGAGAAAGCGGCGCGGAGATG 540
Db 20648 AAACCCGGGTGGCATCTCTCGCGGTGACCCGAGTCTGTAACGAGAAAGCGGCGCGGAGATG 20707
QY 541 CCGGTGTCCGGGGAACAGTGGCGGACCGGCTTCGGATCTGCTGGCTGAGCACTTCTG 600
Db 20708 CCGGTGTCCGGGGAACAGTGGCGGACCGGCTTCGGATCTGCTGGCTGAGCACTTCTG 20767
QY 601 GAATGCTGCTGGACGAGGCGGCTTCGAGATCTTCGATGGGAGAGACGCTGTGTCGAGAGACC 660
Db 20768 GAATGCTGCTGGACGAGGCGGCTTCGAGATCTTCGATGGGAGAGACGCTGTGTCGAGAGACC 20827
QY 661 CGGTACTTATGATCCGAGATTCGCGGAGAGCTCGCTGGCGACACAGACGGGATCCGGAC 720
Db 20828 CGGTACTTATGATCCGAGATTCGCGGAGAGCTCGCTGGCGACACAGACGGGATCCGGAC 20887
QY 721 AGGTACGGGCGGCTGTGCGCGGCTGGGCGCGCGGCTGCGATTATGAAATATATGCC 780
Db 20888 AGGTACGGGCGGCTGTGCGCGGCTGGGCGCGCGGCTGCGATTATGAAATATATGCC 20947
QY 781 CACGACATGGGCTATGCGATTCTGACGCGCGGAGAAAGCGGTCGAGTGA 828
Db 20948 CACGACATGGGCTATGCGATTCTGACGCGCGGAGAAAGCGGTCGAGTGA 20995

RESULT 2
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Balitz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
```

```
; CURRENT APPLICATION NUMBER: US/09/370,700
; FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Query Match      100.0%; Score 828; DB 3; Length 80161;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGTGGCCAGGTGGCCGCAACCATGACGAGGTGGGGAATGATGACCTGGTCAG 60
Db 20168 GTTTCACAGGTGGCCGCAACCATGACGAGGTGGGGAATGATGACCTGGTCAG 20227
QY 61 CCGTGTCTGAATCTGGTGGGCGGCGCCCTGGCCATCCACGAGCTTACTGGAGAAC 120
Db 20228 CCGTGTCTGAATCTGGTGGGCGGCGCCCTGGCCATCCACGAGCTTACTGGAGAAC 20287
QY 121 GACGGGCGGGCTTCTGTCGACGAGGCGCGCCGACCGGCTACCGACCTTGTCCGAAACG 180
Db 20288 GACGGGCGGGCTTCTGTCGACGAGGCGCGCCGACCGGCTACCGACCTTGTCCGAAACG 20347
QY 181 ACCGTGCTGATGAGCGGCGCTGACTGCTGATGATGGGGTGGGCTACCGACCAACGAGCG 240
Db 20348 ACCGTGCTGATGAGCGGCGCTGACTGCTGATGATGGGGTGGGCTACCGACCAACGAGCG 20407
QY 241 CTGGCGGTGCGCGCGACCAACCGGATCCAGATCCAGGATCAACCGTCAAGCGAGTGCAA 300
Db 20408 CTGGCGGTGCGCGCGACCAACCGGATCCAGATCCAGGATCAACCGTCAAGCGAGTGCAA 20467
QY 301 GTGGCCATGCGCGCTGATGTCGCGGACCGGAACTAAGCAACCGGCTGAGACTTCTG 360
Db 20468 GTGGCCATGCGCGCTGATGTCGCGGACCGGAACTAAGCAACCGGCTGAGACTTCTG 20527
QY 361 TCGGTGATGATGATGCTCCCTGCGATCCGACCAATGCTTTGACGCGCGCTGGGCGCATG 420
Db 20528 TCGGTGATGATGATGCTCCCTGCGATCCGACCAATGCTTTGACGCGCGCTGGGCGCATG 20587
QY 421 CAGTGCCTGTGGAGATGTCGGAACCGGACCGGTCCATCCGGAATCTCTTGAAGTACTC 480
Db 20588 CAGTGCCTGTGGAGATGTCGGAACCGGACCGGTCCATCCGGAATCTCTTGAAGTACTC 20647
QY 481 AAACCCGGGTGGCATCTCTCGCGGTGACCCGAGTCTGTAACGAGAAAGCGGCGCGGAGATG 540
Db 20648 AAACCCGGGTGGCATCTCTCGCGGTGACCCGAGTCTGTAACGAGAAAGCGGCGCGGAGATG 20707
QY 541 CCGGTGTCCGGGGAACAGTGGCGGACCGGCTTCGGATCTGCTGGCTGAGCACTTCTG 600
Db 20708 CCGGTGTCCGGGGAACAGTGGCGGACCGGCTTCGGATCTGCTGGCTGAGCACTTCTG 20767
QY 601 GAATGCTGCTGGACGAGGCGGCTTCGAGATCTTCGATGGGAGAGACGCTGTGTCGAGAGACC 660
Db 20768 GAATGCTGCTGGACGAGGCGGCTTCGAGATCTTCGATGGGAGAGACGCTGTGTCGAGAGACC 20827
QY 661 CGGTACTTATGATCCGAGATTCGCGGAGAGCTCGCTGGCGACACAGACGGGATCCGGAC 720
Db 20828 CGGTACTTATGATCCGAGATTCGCGGAGAGCTCGCTGGCGACACAGACGGGATCCGGAC 20887
QY 721 AGGTACGGGCGGCTGTGCGCGGCTGGGCGCGCGGCTGCGATTATGAAATATATGCC 780
Db 20888 AGGTACGGGCGGCTGTGCGCGGCTGGGCGCGCGGCTGCGATTATGAAATATATGCC 20947
QY 781 CACGACATGGGCTATGCGATTCTGACGCGCGGAGAAAGCGGTCGAGTGA 828
Db 20948 CACGACATGGGCTATGCGATTCTGACGCGCGGAGAAAGCGGTCGAGTGA 20995
```

```
RESULT 3
US-09-603-207-1
; Sequence 1, Application US/09603207B
; Patent No. 6521406
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 Div1
; CURRENT APPLICATION NUMBER: US/09/603,207B
; EARLIER FILING DATE: 2000-06-23
; EARLIER APPLICATION NUMBER: 09/370,700
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolypora spinoza
US-09-603-207-1

Query Match      100.0%; Score 828; DB 4; Length 80161;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTTCGACAGTGGCCGACCAACATCCACAGATTTGGGAGATGTATGACTGTGACG 60
DB 20168 GTGTTCGACAGTGGCCGACCAACATCCACAGATTTGGGAGATGTATGACTGTGACG 20227
QY 61 CCGTTGCTGAAGTGGTGGCGGGGGGCGCCCTGCGCATCCACACAGGCTACTGGAGAAC 120
DB 20228 CCGTTGCTGAAGTGGTGGCGGGGGGCGCCCTGCGCATCCACACAGGCTACTGGAGAAC 20287
QY 121 GACGGGGGGGCTTCTTGGGACAGAGCCCGGCTCACCGACTTGTGCGCGAAACG 180
DB 20288 GACGGGGGGGCTTCTTGGGACAGAGCCCGGCTCACCGACTTGTGCGCGAAACG 20347
QY 181 ACCGTGCTCGATGGCGCGGCTTCCATCTGCTGATGTGGGGTGGCGGTACCGAACACGCG 240
DB 20348 ACCGTGCTCGATGGCGCGGCTTCCATCTGCTGATGTGGGGTGGCGGTACCGAACACGCG 20407
QY 241 CTGGCGGTGGGGGGGCGAACAGCGATCCAGATCAACCGGATCAACGCTCAGCGAGGTCAA 300
DB 20408 CTGGCGGTGGGGGGGCGAACAGCGATCCAGATCAACCGGATCAACGCTCAGCGAGGTCAA 20467
QY 301 GTGGCCATTCGCGCGCTGATTTGGGCAAGCGGAAACGCGGACTTAAGCAACCGGGTGGACTTCTCG 360
DB 20468 GTGGCCATTCGCGCGCTGATTTGGGCAAGCGGAAACGCGGACTTAAGCAACCGGGTGGACTTCTCG 20527
QY 361 TGCGTTCGATGTCATGTCCTTCGCTTCCATCCCGGACAAATCTTTCGACGCGCTGGGGCCATG 420
DB 20528 TGCGTTCGATGTCATGTCCTTCGCTTCCATCCCGGACAAATCTTTCGACGCGCTGGGGCCATG 20587
QY 421 CAGTTCGCTGTGAGAGATTCGGAACCGGACCGGTGCATCCCGGAAATCTTTCGAGTACTC 480
DB 20588 CAGTTCGCTGTGAGAGATTCGGAACCGGACCGGTGCATCCCGGAAATCTTTCGAGTACTC 20647
QY 481 AAACCCGGTGGCATCTTCGGGGGTCAACGAGGTGCTCAACGAGAAAGCGGGCGGGGATG 540
DB 20648 AAACCCGGTGGCATCTTCGGGGGTCAACGAGGTGCTCAACGAGAAAGCGGGCGGGGATG 20707
QY 541 CCGGTTCGCGGGGACAGTTCGCGGACCGGGCTTTCGATTCGCTGAGCAACTCTG 600
DB 20708 CCGGTTCGCGGGGACAGTTCGCGGACCGGGCTTTCGATTCGCTGAGCAACTCTCTG 20767
QY 601 GAATTCGCTGCTGACGCGGGGTTGAGATCTTCGATTTGGGAGAGAGTGTCTGAGAGACC 660
DB 20768 GAATTCGCTGCTGACGCGGGGTTGAGATCTTCGATTTGGGAGAGAGTGTCTGAGAGACC 20827
```

```
QY 661 CCGTACTTCATGCGCGAGTTGCGCCGAAGAGCTCGTGGGCAACAGACGCGGATCCGGAC 720
DB 20828 CCGTACTTCATGCGCGAGTTGCGCCGAAGAGCTCGTGGGCAACAGACGCGGATCCGGAC 20887
QY 721 AGGTACGGGCGGAGTGTCCGCGGCTGGGCGCGGCTTGGCATTTATGAAATATGCG 780
DB 20888 AGGTACGGGCGGAGTGTCCGCGGCTGGGCGCGGCTTGGCATTTATGAAATATGCG 20947
QY 781 CACGACATGGGCTATGCGATTTCTGACGGGCGCGGAAACCGGTGGCTGA 828
DB 20948 CACGACATGGGCTATGCGATTTCTGACGGGCGCGGAAACCGGTGGCTGA 20995

RESULT 4
US-08-813-940-3/c
; Sequence 3, Application US/08813940
; Patent No. 5834279
; GENERAL INFORMATION:
; APPLICANT: Rubin, Harvey
; APPLICANT: Yang, Jude
; APPLICANT: Avarbock, David
; APPLICANT: Curran, Sean
; TITLE OF INVENTION: Methods of Identifying Compounds that
; TITLE OF INVENTION: Inhibit DNA Synthesis in Mycobacterium Tuberculosis and
; TITLE OF INVENTION: Composition, Reagents and Kits for Performing the Same
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5834279ris
; STREET: One Liberty Place, 46th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/813,940
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,271
; FILING DATE: 07-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: UPAP-0220
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3107 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 207..2909
; US-08-813-940-3

Query Match      2.4%; Score 20; DB 2; Length 3107;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 ACTTCTGTCGTCGATGCC 372
DB 2143 ACTTCTGTCGTCGATGCC 2124
```

```
RESULT 5
US-08-576-626A-2
; Sequence 2, Application US/08576626A
; Patent No. 5998194
; GENERAL INFORMATION:
; APPLICANT: Summers, R.G.
; APPLICANT: Katz, L.
; APPLICANT: Donadio, S.
; APPLICANT: Steaver, M.J.
; TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR
; TITLE OF INVENTION: BIOSYNTHESIS GENES
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/576,626A
; FILING DATE: 21-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; ATTORNEY/AGENT INFORMATION:
; FILING DATE:
; NAME: Diane Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 5857.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 938-1137
; TELEFAX: (847) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8051 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-576-626A-2

Query Match      2.4%; Score 20; DB 2; Length 8051;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      148 GCGACCGGCTCAGCAGCT 167
DB      6799 GCGACCGGCTCAGCAGCT 6818

RESULT 6
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Query Match      2.4%; Score 20; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      353 ACTTCTCGTGGCTGATGCC 372
DB      4077513 ACTTCTCGTGGCTGATGCC 4077532

RESULT 7
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      2.4%; Score 20; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      353 ACTTCTCGTGGCTGATGCC 372
DB      4085333 ACTTCTCGTGGCTGATGCC 4085352

RESULT 8
US-09-198-452A-1
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffla, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(15000)
; OTHER INFORMATION: n=a or c or g or t
```





```
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (750001)..(765000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (765001)..(780000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (780001)..(795000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (795001)..(810000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (810001)..(825000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (825001)..(840000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (840001)..(855000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (855001)..(870000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (870001)..(885000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (885001)..(900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (900001)..(915000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
```

Query Match 2.3%; Score 19; DB 4; Length 1230025;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 CGATTATGAGAAATATGCC 780  
DB 400440 CGATTATGAGAAATATGCC 400458

RESULT 9  
US-09-438-185A-1  
Sequence 1, Application US/09438185A  
Patent No. 6822071  
GENERAL INFORMATION:  
APPLICANT: Stephens, Richard  
APPLICANT: Mitchell, Wayne  
APPLICANT: Kalman, Sue  
APPLICANT: Davis, Ronald  
TITLE OF INVENTION: Chlamydia Pneumoniae Genome Sequence  
FILE REFERENCE: 018941-000411US  
CURRENT APPLICATION NUMBER: US/09/438,185A  
CURRENT FILING DATE: 2002-03-13  
PRIOR APPLICATION NUMBER: US 60/108,279  
PRIOR FILING DATE: 1998-11-12  
PRIOR APPLICATION NUMBER: US 60/128,606  
PRIOR FILING DATE: 1999-04-08  
NUMBER OF SEQ ID NOS: 1074  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 1230230  
TYPE: DNA  
ORGANISM: Chlamydia pneumoniae  
US-09-438-185A-1

Query Match 2.3%; Score 19; DB 4; Length 1230230;  
Best Local Similarity 100.0%; Pred. No. 14;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 762 CGATTATGAGAAATATGCC 780  
DB 390045 CGATTATGAGAAATATGCC 390063

RESULT 10  
US-09-313-294A-3098  
Sequence 3098, Application US/09313294A  
Patent No. 6476212  
GENERAL INFORMATION:  
APPLICANT: Laligudi, Raghunath V.  
APPLICANT: Ito, Laura Y.  
APPLICANT: Sherman, Bradley K.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR  
FILE REFERENCE: PL-0017 US  
CURRENT APPLICATION NUMBER: US/09/313,294A  
CURRENT FILING DATE: 1999-05-14  
NUMBER OF SEQ ID NOS: 7600  
SOFTWARE: PERL Program  
SEQ ID NO 3098  
LENGTH: 266  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: Incyte ID No. 6476212 700610875H1  
US-09-313-294A-3098

Query Match 2.2%; Score 18; DB 4; Length 266;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 643 GACGTGTCGTGAGGACC 660  
DB 204 GACGTGTCGTGAGGACC 221

RESULT 11  
US-09-313-294A-2444  
Sequence 2444, Application US/09313294A  
Patent No. 6476212  
GENERAL INFORMATION:  
APPLICANT: Laligudi, Raghunath V.  
APPLICANT: Ito, Laura Y.  
APPLICANT: Sherman, Bradley K.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR  
FILE REFERENCE: PL-0017 US  
CURRENT APPLICATION NUMBER: US/09/313,294A  
CURRENT FILING DATE: 1999-05-14  
NUMBER OF SEQ ID NOS: 7600  
SOFTWARE: PERL Program  
SEQ ID NO 2444  
LENGTH: 272  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: Incyte ID No. 6476212 70052677H1  
US-09-313-294A-2444

Query Match 2.2%; Score 18; DB 4; Length 272;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 643 GACGTGTCGTGAGGACC 660  
DB 208 GACGTGTCGTGAGGACC 225

RESULT 12  
US-09-621-976-489

```
; Sequence 489, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTE and Encoded Human Proteins.
; FILE REFERENCE: GENSET 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 489
; LENGTH: 282
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..282
; NAME/KEY: misc_feature
; LOCATION: 74
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-489

Query Match      2.2%; Score 18; DB 4; Length 282;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      426 GCGTGTGAGATGTCGGA 443
DB      81 GCTGTTGGAGATGTCGGA 98

RESULT 13
US-09-513-999C-10082/c
; Sequence 10082, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 10082
; LENGTH: 391
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-10082

Query Match      2.2%; Score 18; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      129 GCGTTCCTGGCAGCAGCC 146
DB      75 GGCTTCCTGGCAGCAGCC 58

RESULT 14
US-09-252-991A-1854
; Sequence 1854, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
```

```
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1854
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1854

Query Match      2.2%; Score 18; DB 4; Length 429;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      739 GCCGGCTGGGCGCCGCG 756
DB      159 GCCGGCTGGGCGCCGCG 176

RESULT 15
US-09-252-991A-1154/c
; Sequence 1154, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1154
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1154

Query Match      2.2%; Score 18; DB 4; Length 513;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      137 GGCAGCAGCGCCGCGACC 154
DB      151 GGCAGCAGCGCCGCGACC 134
```

Search completed: August 29, 2005, 19:39:34  
Job time : 245 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 29, 2005, 17:23:47 ; Search time 698 Seconds  
(without alignments)  
7761.717 Million cell updates/sec

Title: US-10-069-353a-7

Perfect score: 828  
Sequence: 1 gctgtccagtggtgcgaccc.....cgcggaagccggtcgctcga 828

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 7331713 seqs, 3271544945 residues

Word size : 0

Total number of hits satisfying chosen parameters: 14663426

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*  
20: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*  
21: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
22: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
23: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:\*  
24: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
25: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
26: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	828	100.0	80161	17	US-10-329-148A-1
2	2.7	849	15	US-10-156-761-925	Sequence 925, Appl
3	2.7	9025608	15	US-10-156-761-1	Sequence 1, Appl
4	2.4	237	11	US-09-987-899-5572	Sequence 5572, Ap
5	2.4	263	11	US-09-987-899-5443	Sequence 5443, Ap
6	2.4	290	11	US-09-987-899-5444	Sequence 5444, Ap
7	2.4	294	11	US-09-987-899-5442	Sequence 5442, Ap

8	2.4	305	11	US-09-987-899-5441	Sequence 5441, Ap
9	2.4	433	11	US-09-987-899-5571	Sequence 5571, Ap
10	2.4	601	20	US-10-425-115-101089	Sequence 101089, A
11	2.4	1164	17	US-10-369-493-33854	Sequence 33854, A
12	2.4	1643	19	US-10-437-963-33720	Sequence 33720, A
13	2.4	1652	18	US-10-425-114-35015	Sequence 35015, A
14	2.4	1689	18	US-10-425-114-18286	Sequence 18286, A
15	2.4	1896	20	US-10-425-115-140615	Sequence 140615, A
16	2.4	1945	19	US-10-437-963-8409	Sequence 8409, Ap
17	2.4	2053	17	US-10-739-930-5026	Sequence 5026, Ap
18	2.4	2805	17	US-10-282-122A-28732	Sequence 28732, A
19	2.4	2985	17	US-10-282-122A-26421	Sequence 26421, A
20	2.3	628	19	US-10-437-963-31900	Sequence 31900, A
21	2.3	870	19	US-10-767-701-12680	Sequence 12680, A
22	2.3	902	20	US-10-425-115-47646	Sequence 47646, A
23	2.3	1182	15	US-10-156-761-2993	Sequence 2993, Ap
24	2.3	1188	21	US-10-492-928A-8	Sequence 8, Appl
25	2.3	1418	21	US-10-492-928A-11	Sequence 11, Appl
26	2.3	1418	21	US-10-492-928A-69	Sequence 69, Appl
27	2.3	1895	20	US-10-425-115-33974	Sequence 33974, A
28	2.3	123025	17	US-10-289-762-1	Sequence 1, Appl
29	2.2	198	19	US-10-437-963-3364	Sequence 3364, Ap
30	2.2	268	9	US-09-923-876-5454	Sequence 5454, Ap
31	2.2	268	10	US-09-923-876-5454	Sequence 5454, Ap
32	2.2	273	9	US-09-923-876-5094	Sequence 5094, Ap
33	2.2	273	10	US-09-923-876-5094	Sequence 5094, Ap
34	2.2	275	9	US-09-923-876-5296	Sequence 5296, Ap
35	2.2	275	10	US-09-923-876-5296	Sequence 5296, Ap
36	2.2	276	9	US-09-923-876-5430	Sequence 5430, Ap
37	2.2	276	9	US-09-923-876-5511	Sequence 5511, Ap
38	2.2	276	10	US-09-923-876-5430	Sequence 5430, Ap
39	2.2	276	10	US-09-923-876-5511	Sequence 5511, Ap
40	2.2	285	9	US-09-923-876-5479	Sequence 5479, Ap
41	2.2	285	10	US-09-923-876-5479	Sequence 5479, Ap
42	2.2	291	20	US-10-425-115-12019	Sequence 12019, A
43	2.2	295	9	US-09-294-093B-2513	Sequence 2513, Ap
44	2.2	295	9	US-09-294-093B-2996	Sequence 2996, Ap
45	2.2	308	9	US-09-294-093B-3598	Sequence 3598, Ap

#### ALIGNMENTS

RESULT 1  
US-10-329-148A-1  
; Sequence 1, Application US/10329148A  
; Publication No. US20040023343A1  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H  
; APPLICANT: Broughton, Mary C  
; APPLICANT: Crawford, Kathryn P  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Treadway, Patli J  
; APPLICANT: Turner, Jan R  
; APPLICANT: Waldron, Clive  
; TITLE OR INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; FILE REFERENCE: 50469 DIV1  
; CURRENT APPLICATION NUMBER: US/10/329,148A  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/603,207B  
; PRIOR FILING DATE: EARLIER FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/370,700  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 80161  
; TYPE: DNA  
; ORGANISM: Saccharopolyspora spinosa  
US-10-329-148A-1  
Query Match 100.0%; Score 828; DB 17; length 80161;  
Best local Similarity 100.0%; Pred. No. 0;



```

; APPLICANT: O Connell, Keith M.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
; FILE REFERENCE: 16517.258
; CURRENT APPLICATION NUMBER: US 09/987,899
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: US 09/262,979
; PRIOR FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: US 60/076,712
; PRIOR FILING DATE: 1998-03-06
; NUMBER OF SEQ ID NOS: 7341
; SEQ ID NO 5572
; LENGTH: 237
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB143-052-Q1-E1-E4
US-09-987-899-5572

Query Match      2.4%; Score 20; DB 11; Length 237;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      492 CATCTCGGCGTCACCGAG 511
DB      192 CATCTCGGCGTCACCGAG 211

RESULT 5
US-09-987-899-5443
; Sequence 5443, Application US/09987899
; Publication No. US20040116682A1
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; APPLICANT: Liu, Jindong
; APPLICANT: Miller, Phillip W.
; APPLICANT: O Connell, Keith M.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
; FILE REFERENCE: 16517.258
; CURRENT APPLICATION NUMBER: US/09/987,899
; CURRENT FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: US 09/262,979
; PRIOR FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: US 60/076,712
; PRIOR FILING DATE: 1998-03-06
; NUMBER OF SEQ ID NOS: 7341
; SEQ ID NO 5443
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700041823H1
US-09-987-899-5443

Query Match      2.4%; Score 20; DB 11; Length 263;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      492 CATCTCGGCGTCACCGAG 511
DB      162 CATCTCGGCGTCACCGAG 181

RESULT 6
US-09-987-899-5444
; Sequence 5444, Application US/09987899
; Publication No. US20040116682A1
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; APPLICANT: Liu, Jindong
; APPLICANT: Miller, Phillip W.
; APPLICANT: O Connell, Keith M.

```

```

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
; FILE REFERENCE: 16517.258
; CURRENT APPLICATION NUMBER: US/09/987,899
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: US 09/262,979
; PRIOR FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: US 60/076,712
; PRIOR FILING DATE: 1998-03-06
; NUMBER OF SEQ ID NOS: 7341
; SEQ ID NO 5444
; LENGTH: 290
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700094321H1
US-09-987-899-5444

Query Match      2.4%; Score 20; DB 11; Length 290;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      492 CATCTCGGCGTCACCGAG 511
DB      240 CATCTCGGCGTCACCGAG 259

RESULT 7
US-09-987-899-5442
; Sequence 5442, Application US/09987899
; Publication No. US20040116682A1
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; APPLICANT: Liu, Jindong
; APPLICANT: Miller, Phillip W.
; APPLICANT: O Connell, Keith M.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
; FILE REFERENCE: 16517.258
; CURRENT APPLICATION NUMBER: US/09/987,899
; CURRENT FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: US 09/262,979
; PRIOR FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: US 60/076,712
; PRIOR FILING DATE: 1998-03-06
; NUMBER OF SEQ ID NOS: 7341
; SEQ ID NO 5442
; LENGTH: 294
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700214146H1
US-09-987-899-5442

Query Match      2.4%; Score 20; DB 11; Length 294;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      492 CATCTCGGCGTCACCGAG 511
DB      249 CATCTCGGCGTCACCGAG 268

RESULT 8
US-09-987-899-5441
; Sequence 5441, Application US/09987899
; Publication No. US20040116682A1
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; APPLICANT: Liu, Jindong
; APPLICANT: Miller, Phillip W.
; APPLICANT: O Connell, Keith M.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated

```

```

; TITLE OF INVENTION: With the Carbon Assimilation Pathway
; FILE REFERENCE: 16517.258
; CURRENT APPLICATION NUMBER: US/09/987,899
; CURRENT FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: US 09/262,979
; PRIOR FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: US 60/076,712
; PRIOR FILING DATE: 1998-03-06
; NUMBER OF SEQ ID NOS: 7341
; SEQ ID NO 5441
; LENGTH: 305
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 70009567A11
US-09-987-899-5441

Query Match          2.4%; Score 20; DB 11; Length 305;
Best Local Similarity 100.0%; Pred.No. 3.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      492 CATCTCGGCGTCACCGAG 511
DB      202 CATCTCGGCGTCACCGAG 221

RESULT 9
US-09-987-899-5571
; Sequence 5571, Application US/09987899
; Publication No. US20040116682A1
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine
; APPLICANT: Liu, Jingdong
; APPLICANT: Miller, Phillip W.
; APPLICANT: O Connell, Keith M.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
; FILE REFERENCE: 16517.258
; CURRENT APPLICATION NUMBER: US/09/987,899
; CURRENT FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: US 09/262,979
; PRIOR FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: US 60/076,712
; PRIOR FILING DATE: 1998-03-06
; NUMBER OF SEQ ID NOS: 7341
; SEQ ID NO 5571
; LENGTH: 433
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3079-007-Q1-X1-B3
US-09-987-899-5571

Query Match          2.4%; Score 20; DB 11; Length 433;
Best Local Similarity 100.0%; Pred.No. 3.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      492 CATCTCGGCGTCACCGAG 511
DB      237 CATCTCGGCGTCACCGAG 256

RESULT 10
US-10-425-115-101089
; Sequence 101089, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovallic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
```

```

; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 101089
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_23706C.1
US-10-425-115-101089

Query Match          2.4%; Score 20; DB 20; Length 601;
Best Local Similarity 100.0%; Pred.No. 3.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      363 CGTCATGCCATGTCCTGCG 382
DB      481 CGTCATGCCATGTCCTGCG 500

RESULT 11
US-10-369-493-33854
; Sequence 33854, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xiandeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 33854
; LENGTH: 1164
; TYPE: DNA
; ORGANISM: magnetic-containing magnetic coccus
US-10-369-493-33854

Query Match          2.4%; Score 20; DB 17; Length 1164;
Best Local Similarity 100.0%; Pred.No. 3.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      92 GCGCATCCACACGCGCTAC 111
DB      14 GCGCATCCACACGCGCTAC 33

RESULT 12
US-10-437-963-33720
; Sequence 33720, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovallic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
```



```

; SEQ ID NO 33720
; LENGTH: 1643
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_37804C.1
US-10-437-963-33720

Query Match
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2.4%; Score 20; DB 19; Length 1643;

QY 492 CATCTCGGCGTCACCGAG 511
DB 135 CATCTCGGCGTCACCGAG 154

RESULT 13
US-10-425-114-35015
; Sequence 35015, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 35015
; LENGTH: 1652
; TYPE: DNA
; ORGANISM: Zea mays subsp. mexicana
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLTEOSINTE008F02_FLI
US-10-425-114-35015

Query Match
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2.4%; Score 20; DB 18; Length 1652;

QY 492 CATCTCGGCGTCACCGAG 511
DB 233 CATCTCGGCGTCACCGAG 252

RESULT 14
US-10-425-114-18286
; Sequence 18286, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 18286
; LENGTH: 1689
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
```

```

; OTHER INFORMATION: Clone ID: LIB3079-016-D7_FLI
US-10-425-114-18286

Query Match
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2.4%; Score 20; DB 18; Length 1689;

QY 492 CATCTCGGCGTCACCGAG 511
DB 260 CATCTCGGCGTCACCGAG 279

RESULT 15
US-10-425-115-140615
; Sequence 140615, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 140615
; LENGTH: 1896
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_59724C.1
US-10-425-115-140615

Query Match
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2.4%; Score 20; DB 20; Length 1896;

QY 492 CATCTCGGCGTCACCGAG 511
DB 361 CATCTCGGCGTCACCGAG 380

Search completed: August 29, 2005, 19:51:16
Job time : 716 secs
```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 29, 2005, 18:38:52 ; Search time 4680 Seconds  
(without alignments)  
2847.262 Million cell updates/sec

Title: US-10-069-353a-8

Perfect score: 1444  
Sequence: 1 VLPFGAPTSQVQGMVDVLT.....YKRYADMGMALITAKPVG 275

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O/cn2 1/USPTO.spool/US10069353/runac.26082005.172053.14666/app.query.fasta\_1.455  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LDOFCL=0 -LDOPEXT=0  
-UNITS=bits8 -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-LOCAL=GN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10069353 @CGN 1.1 3731 @runac.26082005.172053.14666 -NCPU=6 -ICPU=3  
-NO MMP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*

1: gb ha:\*  
2: gb htg:\*  
3: gb ln:\*  
4: gb om:\*  
5: gb ov:\*  
6: gb pac:\*  
7: gb ph:\*  
8: gb pl:\*  
9: gb pr:\*  
10: gb ro:\*  
11: gb stb:\*  
12: gb sy:\*  
13: gb un:\*  
14: gb vi:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1444	100.0	828	6	AX089422 Sequence
2	1444	100.0	45624	6	AX089419 Sequence
3	1444	100.0	50000	6	AX089416 Sequence
4	1444	100.0	80161	1	AY007564 Saccharop

	5	1444	100.0	80161	6	AR165018
	6	1444	100.0	80161	6	BD137649 Biosynthe
	7	1444	100.0	80161	6	AR281866 Sequence
	8	1315	91.1	36538	6	AX600587 Sequence
	9	441.5	30.6	5998	1	AB032524 Streptom
	10	441.5	30.6	302675	1	AP005024 Streptom
	11	439.5	30.4	36538	6	AX600587 Sequence
	12	439	30.4	1410	6	E07846 DNA sequence
	13	438.5	30.4	846	6	AX598613 Sequence
	14	438.5	30.4	52101	6	AX598593 Sequence
	15	435.5	30.2	900	6	AX089434 Sequence
	16	435.5	30.2	45624	6	AX089419 Sequence
	17	435.5	30.2	50000	6	AX089416 Sequence
	18	435.5	30.2	80161	1	AY007564 Saccharop
	19	435.5	30.2	80161	6	AR165018 Sequence
	20	435.5	30.2	80161	6	BD137649 Biosynthe
	21	435.5	27.8	80161	6	AR281866 Sequence
	22	401.5	26.2	109528	1	AF040570 Amycolato
	23	379	26.2	6085	6	AF0304
	24	379	26.2	6085	6	AR144763 Sequence
	25	376.5	26.1	852	6	BD269407 Mitomycin
	26	376.5	26.1	852	6	AR266893 Sequence
	27	376.5	26.1	53500	6	BD269447 Mitomycin
	28	376.5	26.1	53500	6	AR266933 Sequence
	29	376.5	26.1	63724	1	AF127374 Streptom
	30	375.5	26.0	17512	1	AB071405 Lechevall
	31	375.5	26.0	25681	1	SAR614559 Saccharot
	32	375.5	26.0	26144	1	AB090952 Lechevall
	33	375.5	26.0	28654	1	AF534707 Lechevall
	34	373.5	25.9	22999	1	AB088119 Streptom
	35	372.5	25.8	846	1	AY396042 Streptom
	36	361.5	25.0	30000	6	AX250261 Sequence
	37	361.5	25.0	103450	1	AF440781 Streptom
	38	345	23.9	828	6	BD269406 Mitomycin
	39	345	23.9	828	6	AR266892 Sequence
	40	343.5	23.8	12381	6	BD217469 Polyclid
	41	343.5	23.8	12381	6	AX006889 Sequence
	42	340	23.5	12637	1	AF323753 Streptom
	43	329.5	22.8	21064	1	AB088224 Streptom
	44	320.5	22.2	2122	6	AF0301 Sequence
	45	320.5	22.2	2122	6	AR144762 Sequence

#### ALIGNMENTS

RESULT 1	AX089422	828 bp	DNA	linear	PAT 21-MAR-2001
LOCUS	AX089422	Sequence 7 from Patent WO0116303.			
DEFINITION	AX089422				
ACCESSION	AX089422.1	GI:13443683			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source					
1..828					
/organism="Saccharopolyspora spinosa"					
/mol_type="unassigned DNA"					
/db_xref="taxon:60894"					
1..828					
/note="unnamed protein product; ORF1: O-Methyltransferase"					
/codon_start=1					
/transl_table=11					
/protein_id="CAC34874.1"					
/db_xref="GI:13443684"					

/translation="MLPGAPTSQGVQWYDVLTPRLNSVAGSPCAIHGWENDGARA  
SWQCAADRLTDLVAERTVLDGCVGRLDVCCTGGOPALRAVANAIOIGITVSQVOVA  
IADQADREGLSHRVDFSCVDAMSLPYPNNAIDPAMAMSLMESEPDALIELRLV  
KRGILGTEVVKREAGGMPVSGDRWPGIARICIAEOLLESLRAAGEPILMEDEVSS  
RTRVEMPQAEELAAHQHIDIRYBPAAVAGMAAACVDYEKVAHMDNGYALITRKRVG"

## ORIGIN

Alignment Scores:  
Pred. No.: 1,86e-100 Length: 828  
Score: 1444.00 Matches: 275  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-069-353a-8 (1-275) x AX089422 (1-828)

QY 1 ValLeuProGlyValAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20  
DB 1 GTGTTGCCAGGTGGCGCACCAATCGCAGAGTTGGGCAATGATGACCTGGTCAAG 60  
QY 21 ProLeuLeuAsnSerValAlaGlyProCysAlaIleHisGlyTyrTrpGluAsn 40  
DB 61 CCGTTGCTAACTCGGTCCGGCGGAGCCCTTGCCATCCACAGGCTACTGGAGAAC 120  
QY 41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60  
DB 121 GACGGGCGGGCTTCCGTGGCAGCAGGCGCGCACCGGCTCACCGACTTGTCCCGAACG 180  
QY 61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80  
DB 181 ACCGTGCTCGATGGCGGCTTCGACTGCTGATGTGGGGTGGCGTACCGCAACACAGCG 240  
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100  
DB 241 CTGCGGCTCGCGGACACCGCATCCAGATCACCGGATCACCGTCAAGCAGGTGCA 300  
QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120  
DB 301 GTGGCCATGGCGCGTGGATTGGCGCAGCAGGAGCGGACTAGGCACTTCTCG 360  
QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140  
DB 361 TCCGTGCAATGCATGCTCCGTGCAACCGGCAATGCTTTCGACGCCCTGGGCGCATG 420  
QY 141 GlnSerLeuLeuGlnMetSerGlnProAspArgAlaIleArgGluIleLeuArgValLeu 160  
DB 421 CAGTGCCTTGGAGATGTCGAACCGGACCGTGCATCCGGAAATCTTTCGAGTACTC 480  
QY 161 LysProGlyGlyIleLeuGlyValThrGluValValIysArgGluAlaGlyGlyMet 180  
DB 481 AAACCCGGGTGCATCTCGGCGTCAACGAGTCTGCAACGAGAGCGGGCGGCGGATG 540  
QY 181 ProValSerGlyAspArgTrpProThrArgIleLeuArgIleCysLeuAlaGluGlnLeu 200  
DB 541 CCGGAGTCCGAGGACAGGTGGCGGACCGGCTTCGATTCGCTCGAGCAACTTCTCG 600  
QY 201 GlnSerLeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220  
DB 601 GAATGCTGCTGCAGCGGGGTTCCAGATCTCGATTGGAGAGAGCTGCTCGAGGACC 660  
QY 221 ArgTyrPheMetProGlnPheAlaGluGluLeuAlaAlaHisGlyIleGlyIleAlaAsp 240  
DB 661 CCGTACTTCATGCGCAGAGTTCGCCGAAAGAGTCTGCTGGCGCACAGAGCGGATCGGGAC 720  
QY 241 ArgTyrGlyProAlaValAlaGlyTyrTrpAlaAlaAlaValCysAspTyrGlyIleGlyIle 260  
DB 721 AGTACCGGGCCCGCTGTCCCGCTGGGCGCGCGGCTTCGATTAAGAAATATATGTC 780  
QY 261 HisAspMetGlyTyrAlaIleLeuThrAlaArgLysProValGly 275  
DB 781 CACGACATGGGCTATGCGATCTCGAGCGGCGGAAAGCCGCTGCGC 825

## RESULT 2

AX089419/c

LOCUS

Sequence 4 from Patent WO0116303.

DEFINITION

AX089419

ACCESSION

AX089419.1

VERSION

GI:13443680

KEYWORDS

SOURCE

ORGANISM

Saccharopolyspora spinosa

Saccharopolyspora spinosa

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.

REFERENCE

Eberz, G., Moehrle, V., Froede, R., Velten, R. and Salas, J.A.

Nucleic acids which code for the enzyme activities of the spinosyn

biosynthesis

JOURNAL

Patent: WO 0116303-A 4 08-MAR-2001;

BAYER AG (DE)

FEATURES

source

1. 45624

/organism="Saccharopolyspora spinosa"

/mol\_type="unassigned DNA"

/db\_xref="taxon:60894"

## ORIGIN

Alignment Scores:  
Pred. No.: 1.57e-98 Length: 45624  
Score: 1444.00 Matches: 275  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-069-353a-8 (1-275) x AX089419 (1-45624)

QY 1 ValLeuProGlyValAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20  
DB 828 GTGTTGCCAGGTGGCGCACCAATCGCAGAGTTGGGCAATGATGACCTGGTCAAG 769  
QY 21 ProLeuLeuAsnSerValAlaGlyProCysAlaIleHisGlyTyrTrpGluAsn 40  
DB 768 CCGTTGCTAACTCGGTCCGGCGGAGCCCTTGCCATCCACAGGCTACTGGAGAAC 709  
QY 41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60  
DB 708 GACGGGCGGGCTTCCGTGGCAGCAGGCGCGCACCGGCTCACCGACTTGTGCGGAAAG 649  
QY 61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80  
DB 648 ACCGTGCTCGATGGCGGCTTCGACTGCTGATGTGGGGTGGCGTACCGCAACACAGCG 589  
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100  
DB 588 CTGCGGCTCGCGGCGGACACCGCATCCAGATCCAGGCACTCACCGTCAAGCAGGTGCA 529  
QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120  
DB 528 GTGGCCATGGCGCGTGGATTGGCGCAGCAGGAGCTAAACCAACCGGCTGAGCTTCTCG 469  
QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140  
DB 468 TCCGTGCAATGCATGCTCCGTCCGACCGGCAATGCTTTCGAGCGCGCTGGGCGCATG 409  
QY 141 GlnSerLeuLeuGlnMetSerGlnProAspArgAlaIleArgGluIleLeuArgValLeu 160  
DB 408 CAGTGCCTTGGAGATGTCGAACCGGACCGTGCATCCGGAAATCTTTCGAGTACTC 349  
QY 161 LysProGlyGlyIleLeuGlyValThrGluValValIysArgGluAlaGlyGlyMet 180  
DB 348 AAACCCGGGTGCATCTCGGCGTCAACGAGTCTGCAACGAGAGCGGGCGGCGGATG 289  
QY 181 ProValSerGlyAspArgTrpProThrArgIleLeuArgIleCysLeuAlaGluGlnLeu 200  
DB 288 CCGGTGTCGGGAGACAGGTGGCGACCGGCTTCGATCTGCTGGCTGAGCAACTTCTG 229

QY 201 GUSerLeuArgAlaIaIaIaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220  
 DB 228 GAATCGCTCGTGCACCGGGGTTTCGAGATCTCGATTGGAGGACGGTGTGAGAGACC 169  
 QY 221 ArgTyrPheMetProGlnPheIaIaGluIleLeuAlaIaIaGlnIaIaIaAsp 240  
 DB 168 CGGTACTTTCATGCCGAGTTCCGCGAAGACTTCGTCGCGACCAAGACGGATCCGGGAC 109  
 QY 241 ArgTyrGlyProAlaValaIaIaGlyTyrPheAlaIaIaValaIaIaAspTyrGluIleVala 260  
 DB 108 AGGTACGGGCGGCTGTGCGCGGCTGGCGCGGCTGCGATTATGAGAAATATATGCC 49  
 QY 261 HisAspMetGlyTyrAlaIaIaIleuThrAlaArgLysProValGly 275  
 DB 48 CACGACATGGGCTATGCGATTCTGACGGCGCGGAGCCGGTCCGC 4  
 RESULT 3  
 AX089416 AX089416 50000 bp DNA linear PAT 21-MAR-2001  
 DEFINITION Sequence 1 from Patent WO0116303.  
 ACCESSION AX089416  
 VERSION AX089416.1 GI:13443677  
 KEYWORDS  
 SOURCE Saccharopolyspora spinosa  
 ORGANISM Saccharopolyspora spinosa  
 Bacteria; Actinobacteridae; Actinomycetales;  
 Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.  
 REFERENCE  
 AUTHORS Berr, G., Moehle, V., Froede, R., Velten, R. and Salas, J. A.  
 TITLE Nucleic acids which code for the enzyme activities of the spinosyn  
 biosynthesis  
 JOURNAL Patent: WO 0116303-A 1 08-MAR-2001;  
 BAYER AG (DE)  
 FEATURES  
 source 1..50000  
 location/Qualifiers  
 /organism="Saccharopolyspora spinosa"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:60894"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.74e-98 Length: 50000  
 Score: 1444.00 Matches: 275  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0  
 US-10-069-353a-8 (1-275) x AX089416 (1-50000)  
 QY 1 ValLeuProGlyGlyAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20  
 DB 44797 GTGTGTCAGTGGCGGACCAACATCCACAGGTTGGGAGAGTATGACTGTGTACG 44856  
 QY 21 ProLeuLeuAsnSerValaIaIaGlyGlyProCysAlaIleHisGlyTyrTrpGluAsn 40  
 DB 44857 CCGTTGCTGAACCTGCTGCGGCGGCGGCCCTGCGCCATCCACACCGGCTACTGGAGAAC 44916  
 QY 41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValaIaIaGluArg 60  
 DB 44917 GACCGGCGGCTTCTCTGCGACGACGCGCGACCGGCTCACCGACTTGTGCGGAGCG 44976  
 QY 61 ThrValLeuAspGlyGlyValaIaIaArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80  
 DB 44977 ACCGTGCTCATGGCGGCTTTCGACTGCTGATGTGGGGTGGCGGTACCGAGCAACGCG 45036  
 QY 81 LeuArgValaIaIaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100  
 DB 45037 CTGCGCGTCCGCGCGCAACGCGATCCAGATCCCGCAACACCGTACGCGACGAGTGC 45096  
 QY 101 ValaIaIaIaIaIaAspCysAlaArgGluArgGlyLeuSerHisArgValaIaIaPheSer 120

DB 45097 GTGGCCATGCCGCTGATTGCCACGCGAACCAGCTAAGCCACCGGTTGACTTCCG 45156  
 QY 121 CysValaAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaIaIaTrpAlaMet 140  
 DB 45157 TCGTGTGATGCCATGTCCTGCGCGTACCGGACCAATGTTTGACGCCGCTTGGGCCATG 45216  
 QY 141 GlnSerLeuLeuGlnMetSerGlnProAspArgAlaIleArgGluIleLeuArgValaIleu 160  
 DB 45217 CAGTCCCTTGTGAGATATGTCGAACCGGACCGTGCATCCGGGAATCTTCGAGTATTC 45276  
 QY 161 LysProGlyGlyIleLeuGlyValaIaIaIleuValaIaIaValaIaIaGlyValaIaIaMet 180  
 DB 45277 AAACCCGATGATCTTCGCGCTCACCGAGTGTGCAACGAGACCGGCGCGGAGTGC 45336  
 QY 181 ProValSerGlyAspArgTrpProThrArgLysLeuArgIleCysLeuAlaGluIleLeuLeu 200  
 DB 45337 CCGGTGTCGGGAGACAGTGGCGGACCGGCTTCGATGTGCTGCTAGCACTTCTTG 45396  
 QY 201 GUSerLeuArgAlaIaIaIaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220  
 DB 45397 GAATCGCTCGTGCACCGGGGTTTCGAGATCTCGATTGGAGAGACGTGTGTCGAGGAC 45456  
 QY 221 ArgTyrPheMetProGlnPheIaIaGluIleLeuAlaIaIaGlnIaIaIaAsp 240  
 DB 45457 CGGTACTTTCATGCGGAGTTCCGCGAAGACTTCGTCGCGACACGACGCGGATCCGCGAC 45516  
 QY 241 ArgTyrGlyProAlaValaIaIaGlyTyrPheAlaIaIaValaIaIaAspTyrGluIleVala 260  
 DB 45517 AGGTACGGGCGGCTGTGCGCGGCTGGCGCGGCTGCGATTATGAGAAATATATGCC 45576  
 QY 261 HisAspMetGlyTyrAlaIaIaIleuThrAlaArgLysProValGly 275  
 DB 45577 CACGACATGGGCTATGCGATTCTGACGGCGCGGAGCCGGTCCGC 45621  
 RESULT 4  
 AY007564 80161 bp DNA linear BCT 16-MAY-2001  
 LOCUS AY007564  
 DEFINITION Saccharopolyspora spinosa probable lyxR-type transcriptional  
 regulator, probable keto acyl reductase, spinosad biosynthetic gene  
 cluster, complete sequence, and probable exodeoxyribonuclease V  
 genes, complete cds; and unknown gene.  
 ACCESSION AY007564.1 GI:13162633  
 VERSION  
 KEYWORDS  
 SOURCE Saccharopolyspora spinosa  
 ORGANISM Saccharopolyspora spinosa  
 Bacteria; Actinobacteridae; Actinomycetales;  
 Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.  
 REFERENCE  
 AUTHORS Waldron, C., Matsushima, P., Rosteck, P.R. Jr., Broughton, M.C.,  
 Turner, D., Madduri, K., Crawford, K.P., Merlo, D.J. and Baltz, R.H.  
 TITLE Cloning and analysis of the spinosad biosynthetic gene cluster of  
 Saccharopolyspora spinosa  
 JOURNAL Chem. Biol. 8 (5), 487-499 (2001)  
 MEDLINE 21257765  
 PUBMED 11358695  
 REFERENCES  
 1 (bases 1 to 80161)  
 2 (bases 1 to 80161)  
 Waldron, C., Matsushima, P., Rosteck, P.R. Jr., Broughton, M.C.,  
 Turner, D., Madduri, K., Crawford, K.P., Merlo, D.J. and Baltz, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-AUG-2000) Dow Agrosciences, 9330 Zionsville Rd.,  
 Indianapolis, IN 46268, USA  
 FEATURES  
 source 1..80161  
 location/Qualifiers  
 /organism="Saccharopolyspora spinosa"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:60894"  
 complement(1135..1971)  
 /note="ORF L16"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable lyxR-type transcriptional regulator"  
 /protein\_id="AAG23282.1"

gene  
/db\_xref="GI:13162654"  
/translation="METRELYFVAVAEELHFGRAAORLGIQAQPLSRITIAQLEORLG  
VLLQRTSRKSLTREGAMLLTEGALIGALRRTORRATISQPSLVIAKAGASG  
ELLAKLIDVAAEPGAVAVDILICSOQKTHDRADYALIHOPDPPTAEIDILIN  
TEQOYAILPTSHPLASERHVRADVSSIPDLPLAMPBGDGYPDGREGVVRNQQLP  
QMIALGRTTVVMBESSRVNLEGLAAVFLVADVTVTIAMPHRSRSLALGLVAVAT  
IL"  
2024. .2791  
/note="ORF 115"  
/codon\_start=1  
/transl\_table=11  
/product="probable keto acyl reductase"  
/protein\_id="AAG23281.1"  
/db\_xref="GI:13162653"  
/translation="WLVRRPRRTLESMEQOTIALVTGANKGIEYIAAGALGWS  
VIGARDHORGEDAVAKLRADGDVAFVSLDVTDOASVAAAAALEEARGLDVYVN  
AGIAGWPEEPSTVPASIRAVETNVIGVAVTNAM.PILPRSRPRITVNSSHVAS  
LTLQTTPEVDLGGISGAYSPSKTFINATITIOYAKLSDTNINKINACGCVATDLNGF  
HQTSTPADGARAIIRLATLPDDGPTGGMFDGANVPW"  
complement(3416. .4165)  
/gene="spnS"  
complement(3416. .4165)  
/note="spnS"  
/gene="spnS"  
/note="probably involved in forosamine biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="probable N-dimethyltransferase"  
/protein\_id="AAG23280.1"  
/db\_xref="GI:13162652"  
/translation="MSRVSDTFEATSSVYSPDHADIYAIHSARGDWAAGEVYQL  
VTRLPEAOSLIDVACGTGAHLERPRAEAKVAGLESDAMKEIIRIVPEVPIHIGD  
IRDFLDGPFVDITCLCTFAAYRTVDLIRVTNMMARHAPGVAVLEPWFPKFI  
DOFVGAVAHGERIVISRLSHSELEGRSRMVRVTVAAPEGIRDFTEILLSFTED  
EXTALPDAGIRAEYLPGAPNGRGLFVGIRN"  
complement(4168. .5325)  
/gene="spnR"  
complement(4168. .5325)  
/note="spnR"  
/gene="spnR"  
/note="probably involved in forosamine biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="probable aminotransferase"  
/protein\_id="AAG23279.1"  
/db\_xref="GI:13162651"  
/translation="MINLHQPLITGEELDAIEVFAASWIGIGPRTTPFAEFAHILG  
VDPEVFLNPGSTAAFLITVOYLDIGREDVTLPSISVAAANAIASSGARPFCVDV  
PRTLPRLDDVRAITTPATKAVLLHYGSGPEVVAIDPCKEKLMLIEDSACAVAS  
SVHGTACGTFGDLATWSPDAMKILVTGCGMFYADPELHARARLAYHLEOMSGFD  
SAKSNRMWDIREDIDGRLIGNDTALGVSQRLPEFINRREIATQYDRLISDV  
GVLLPPTLPDGHVSHYFWYQALAEIRDVOAQOMLERGIYSRYAPLKHVPYIRA  
DCKLPASADACGRTLLPLRPBSLDABEVRTVADEFQKAVENHISGRSLRK"  
complement(5363. .6751)  
/gene="spnQ"  
complement(5363. .6751)  
/note="spnQ"  
/gene="spnQ"  
/note="probably involved in forosamine biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="probable NDP-hexose-3,4-dehydratase"  
/protein\_id="AAG23278.1"  
/db\_xref="GI:13162650"  
/translation="MOSRKTALGKGARVYNSCDDTCATTEMVPADKRIILASVRYD  
HREGSPTPVAGSTPIRPSGAVLDEDDRYALVYALERIRIAGNAARFESBFAFPG  
LRKALVNSGSSANILALSLTSPKLGARLRPGSEVITAAVGFPTTINPVAONGLVP  
FVDELGYNATPDIRIKAAVETRTAIMEATLGNPFADEIAMEIAKHEHFLIVEDN  
CDVAGSYRGRITGTFGDLITVSPFAHHITSGEGCVLKTGLIELARIIESLRDGRD  
CWCEGVNTGCRKRFYDHLGLPFGYDHKYTSGSHVGNLKTDLQALALASOLSKISA  
FGSARRMMRRLREGLSGIPGLIPVATPHSPSPSGEATITISADAGFTRALVMELE  
SRNITRLLFGNITRHPAFBOYRIRIDALTNDSITVDRITGVGVPGITQMDMDIV  
VESIAEFVAKSS"

gene  
7083. .8450  
/gene="spnP"  
/note="spnP"  
/note="involved in forosamine addition"  
/codon\_start=1  
/transl\_table=11  
/product="probable NDP-forosamyltransferase"  
/protein\_id="AAG23277.1"  
/db\_xref="GI:13162649"  
/translation="MILGMLPGCSIAIGEFNRVLTPLPASSHFNTVPLAMALRAG  
HEVVAICENNVSWTGAGLTAVPVGELDLISLAAKNEVLGSGVSFDEGRHPELP  
DELSINSGRDVAVEQLHLVDRSLDLDLGFARKWQPDLVWDMAMVCSGVARALG  
ARHVRMLVALDVSGLMRSGFLEYOESKPEBQVDPDLGTWLGAKAKRGAIFDEBIVTG  
QATIDPIPSMWRLLPVLDLYISMRVPVNGPRLVPEMLDERPTKRCVITGGLTRRLS  
RVTEYOESQVSDHMAWBLRLGARADVEVITATLSDDEREMGELSPNVRHVEVPL  
NELSEGSVITHHSTTQETNATNGVQPOLIPCTPMDESRARLADRAAGIYLDPA  
TFTEDDVRGOLARLIDEPSPANALIRREIEBSPPHDIVPRLKLVARENRRTQ  
SDGRP"  
8967. .10427  
/gene="spnO"  
8967. .10427  
/note="spnO"  
/gene="spnO"  
/note="involved in forosamine biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="probable NDP-hexose-2,3-dehydratase"  
/protein\_id="AAG23276.1"  
/db\_xref="GI:13162648"  
/translation="MSSSVEAEASAAAPLGSNNTRRVDSALSNCMITPTERHCL  
ADRIGENSFETNRJPEFRLSKWKFDASTEMLVHADGFPFVEGLQVETNGAASWQ  
PIINQAEVGLIGILVKEIDGVLKMSAKPGVNVNLOQSPVQATRSYNTQAHRS  
VPVPLDFLGRGSRVLDVYLOSQSGWFRKRRNRNVVEVEEVPVLPFCMLTGQ  
VVALRDQNTVMNMTTRVLSGIPPHDSATGTELAASHEPFOAVARLSHGIDSSIS  
EAVGWPEAKARVRLRATRVPLSVDMVWTFDTEIADQDGFAVIVIVSATSATREVA  
SWTPMIEPRQGEIALLVKRIGSVLGHVHARIEAKYKTAELAPVQGSVANTQST  
PSNDMPPLDDVLTADPEVTYESILSEBGRFPYQAOENRRIIEVHEDFARPSPDRT  
WMTLGQGEELRSTHPLNIQARSLVASHSLMALGR"  
10436. .11434  
/gene="spnM"  
10436. .11434  
/note="spnM"  
/gene="spnM"  
/note="involved in forosamine biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="probable NDP-hexose-3-ketoreductase"  
/protein\_id="AAG23275.1"  
/db\_xref="GI:13162647"  
/translation="WRRPVRIGVLCASFAWRMLPAMCDVAETEVAVASRDBAKAE  
RFAARFCEAVLGYORLLEBPDIDAVVPLPPGMAEMIGKALEADKHVLAKEPLTTT  
ASDPAARLVGLARKNLLRENYLLEHHGRDVVDLQSGEIGELRFTAVFGIPLP  
TDTIRYRETELGGGALLDIGVPAARAHFLPLGPTVAGSHASHEOESGVLGSGVLL  
SEGCTVAHIGPCHYHRAVYELMWSGRGTVVPRAPPEAWQAVITIEBKGVYDELS  
LPARDQKAVTAPARDIRAGTCVDDPAVAGDSGSMIQQAALVEAIGQARCGST"  
complement(11530. .12492)  
/gene="spnM"  
complement(11530. .12492)  
/note="spnM"  
/gene="spnM"  
/note="involved in spinosyn aglycone biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="spnM"  
/protein\_id="AAG23274.1"  
/db\_xref="GI:13162646"  
/translation="MPAAVGGTVLVPIPMWRERDPLITFAVNGHGLGSOVAPSYLR  
TGTEPERELIIVALDRGMAYVITDYBGLGPGTHTTYVGAQAGAMDDAARAARLGG  
SGLTITDCPVGIVGAQGGQASAPFGEJLHPTAYELRLRAAAGAVPTDLDLITHRNDG  
VFTCPVAVGLARGLAAVAPDLPEBELTEAGTAVDQVELEIAGAPLVIRFGRELSDFL  
DVTSLFQCPQPTMRGLASVAGRNQGPVPTLVNHSITDDELIVPFAFGELRLDSYRAAGT  
PVRNHPSGLAHPAALASRVVSWDFEHSBSAISGPRDAN"  
complement(12696. .13547)  
/gene="spnL"  
complement(12696. .13547)

```
/gene="spnL"
/Note="Involved in spinosyn aglycone biosynthesis;
possibly involved in polyketide bridging"
/codon_start=1
/transl_table=1
/product="methyltransferase-like protein"
/protein_id="AAG23273.1"
/db_xref="GI:13162645"
/translation="MESTFDALHGRPLHNGWAGYREDACATPMSDAADLTDLFI
DKALAPGAHLFDLGCNGQPVVRAACASGVATGTVNAQHILAAATRLANETLGS
LEFDLVDGAQLPYDGFPOAAMQSVQVVDPAALREVHRLIEBGRFVLDIITR
VRLPEEYAAVWTGTATLNSFPALVSEGEFLEVTDLAOTRCWMSVYDELLRL
DELAGEPAAVGTQQRYIGDIAKKHPPAQILAAVAEYRKHPDVARREBSMFLL
QARKKS"
complement(13592..14785)
/gene="spnK"
complement(13592..14785)

Alignment Scores:
Pred. No.: 2,93e-98 Length: 80161
Score: 1444.00 Matches: 275
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-069-353a-8 (1-275) x AY007564 (1-80161)

QY 1 ValLeuProGlyValAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20
20168 GTGTTGCGAGTGGCGGCGACCAACATCGCAGAGTTGGGAGATGATGACCTGGTACG 20227
QY 21 ProLeuLeuAsnSerValAlaGlyGlyProCyAspAlaIleHisGlyTyrTrpGluAsn 40
20228 CCGTTGCTGAATCGGTGCGGGGCGCCCTGCGCCATCCACACCGGCTACTGGGAGAAC 20287
QY 41 AspGlyValArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
20288 GACGGGGGGGCTTCCTGCGAGAGCGCGGCGGCTGACCGACTTGTGCGCGAAGCGG 20347
QY 61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCyAspGlyThrGlyGlnProAla 80
20348 ACCGTGCTCATGCGCGCGCTTCACTGCTGATGCGGGGTGCGGTACCGGACCAACGCG 20407
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyTyrLeuValSerGlnValGln 100
20408 CTGGCGCTGCGCGCGGCAACGCGATCCAGATCCCGGCAATCACCGTCAAGCGAGTCAA 20467
QY 101 ValAlaIleAlaAlaAspCyAspAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
20468 GTGGCCATCGCCGCTGATGGCGAGCGGAACTGAGCCACCGGGTGGACTTCTCG 20527
QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
20528 TGGCTCATGCTCATGCTCCTGCGCTACCGGACAACTCTTTCAGCGCGCTGGGCCATG 20587
QY 141 GlnSerLeuLeuGlnMetSerGluProAspArgAlaIleArgGlyIleLeuArgValLeu 160
20588 CAGTCGCTGTTGGAGATGTCGGAACCGGACCGTCCATCCGGGAAATCTTTCAGTACTC 20647
QY 161 LysProGlyGlyIleLeuGlyValThrGluValValLysArgGlyAlaGlyGlyGlyMet 180
20648 AAACCCGATGCGATCCGCGGCGTACCGAGGTCCTCAACCAAGAAAGCGGCGGGGATG 20707
QY 181 ProValSerGlyAspArgTrpProThrGlyLeuArgIleCysLeuAlaGlnGlnLeu 200
20708 CCGGTGTCGGGGGACAGGTGGCCACCGGCTTGGATGCTCCGCTGGAGCACTTCTG 20767
QY 201 GluSerLeuArgAlaAlaGlyPheGlnIleLeuAspTrpGluAspValSerSerArgThr 220
20768 GAATCGCTGCGTGCAGCGGGGTTCGAGATCTCGATTGGGAGGACGCTGCTCGAGGACC 20827
QY 221 ArgTyrThrMetProGlnPheAlaGlnGluLeuAlaAlaHisGlnHisGlyIleAlaAsp 240
```

```
DB 20828 CGGTACTTCATGCCGAGTTCGCCGGAAGAGCTCGCTGGCGACCAACGCGGATCCGCGAC 20887
QY 241 ArgTyrGlyProAlaValAlaGlyTyrTrpAlaAlaAlaValCysAspTyrGluArgTyrAla 260
20888 AGGTACGGGCGCGGCTGTCGCGCGGCGCGCGGCTGCGATTAAGAAATATGCGC 20947
QY 261 HisAspMetGlyTyrAlaIleLeuThrAlaGlyLysProValGly 275
DB 20948 CACGACATGGGCTATGCGATTCGACGCGCGGAAACCGGTGCGC 20992

RESULT 5
LOCUS ARI65018 80161 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 1 from patent US 6274350.
ACCESSION ARI65018
VERSION ARI65018.1 GI:16238344
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 80161)
AUTHORS Balcz,R.H., Crawford,K.P., Broughton,M.Christine., Madduri,K.,
Merlo,D.J., Turner,J.R., Treadway,P.J. and Waldron,C.
TITLE Biosynthetic genes for spinosyn insecticide production
JOURNAL Patent: US 6274350-A 1 14-AUG-2001;
FEATURES
Source
1..80161
/organism="unknown"
/mol_type="unassigned DNA"

Alignment Scores:
Pred. No.: 2,93e-98 Length: 80161
Score: 1444.00 Matches: 275
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-069-353a-8 (1-275) x ARI65018 (1-80161)

QY 1 ValLeuProGlyValAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20
20168 GTGTTGCGAGTGGCGGCGACCAACATCGCAGAGTTGGGAGATGATGACCTGGTACG 20227
QY 21 ProLeuLeuAsnSerValAlaGlyGlyProCyAspAlaIleHisGlyTyrTrpGluAsn 40
20228 CCGTTGCTGAATCGGTGCGGGGCGCCCTGCGCCATCCACACCGGCTACTGGGAGAAC 20287
QY 41 AspGlyValArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
20288 GACGGGGGGGCTTCCTGCGAGAGCGCGGCGGCTGACCGACTTGTGCGCGAAGCGG 20347
QY 61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCyAspGlyThrGlyGlnProAla 80
20348 ACCGTGCTCATGCGCGCGCTTCACTGCTGATGCGGGGTGCGGTACCGGACCAACGCG 20407
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyTyrLeuValSerGlnValGln 100
20408 CTGGCGCTGCGCGCGGCAACGCGATCCAGATCCCGGCAATCACCGTCAAGCGAGTCAA 20467
QY 101 ValAlaIleAlaAlaAspCyAspAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
20468 GTGGCCATCGCCGCTGATGGCGAGCGGAACTGAGCCACCGGGTGGACTTCTCG 20527
QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
20528 TGGCTCATGCTCATGCTCCTGCGCTACCGGACAACTCTTTCAGCGCGCTGGGCCATG 20587
QY 141 GlnSerLeuLeuGlnMetSerGluProAspArgAlaIleArgGlyIleLeuArgValLeu 160
20588 CAGTCGCTGTTGGAGATGTCGGAACCGGACCGTCCATCCGGGAAATCTTTCAGTACTC 20647
```

QY 161 LysProGlyValIleuGlyValThrGluValValIleArgGluAlaGlyGlyMet 180  
 Db 20648 AAACCCGGGCGATCTCCGGCTCACCGAGTCGTCAACGAGACGGGGCGGATG 20707  
 QY 181 ProValSerGlyAspArgTrpProThrglyLeuValIleCysLeuAlaGluGlnLeu 200  
 Db 20708 CCGGTGTCCGGGAGCAGGTGGCGGACCGGCTTCGGATCTGCTGGCTAGCACTTCTG 20767  
 QY 201 GluSerLeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220  
 Db 20768 GAATGCTGCGTGCAGCGGGGTTCAGATCTCGATTGGAGACGTCGTGAGAGACC 20827  
 QY 221 ArgTyrPheMetProGlnPheAlaGluGlnLeuAlaAlaHisGlnHisGlyIleAlaAsp 240  
 Db 20828 CGGTACTTCATGCCGAGTTCCCGAAGAGCTCGCTGGCCACACAGACGGGATCCGGAC 20887  
 QY 241 ArgTyrGlyProAlaValAlaGlyTyrTrpAlaAlaAlaValCysAspTyrGluLysTyrAla 260  
 Db 20888 AGGTACGGGCGCGCTGTCCGCGGTGGCGCGCGCGCTGCTCGATTATGAGAAATATGCC 20947  
 QY 261 HisAspMetGlyTyrAlaAlaIleuThrAlaArgLysProValGly 275  
 Db 20948 CACGACATGGGCTATGCGATCTGACGCGCGGAGACCGGTGCGC 20992  
 RESULT 6  
 LOCUS BD137649 80161 bp DNA linear PAT 18-SEP-2002  
 DEFINITION Biosynthetic genes for spinosyn insecticide production.  
 ACCESSION BD137649  
 VERSION BD137649.1 GI:23232594  
 KEYWORDS JP 2002505881-A/1.  
 SOURCE Saccharopolyspora spinosa  
 ORGANISM Saccharopolyspora spinosa  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Pseudonocardiaceae; Pseudonocardiaceae; Saccharopolyspora.  
 1 (bases 1 to 80161)  
 REFERENCE Baltz,R.H., Broughton,M.C., Crawford,K.P., Madduri,K., Merlo,D.J.,  
 Treadway,P.J., Turner,J.R. and Waldron,C.  
 TITLE Biosynthetic genes for spinosyn insecticide production  
 JOURNAL Patent: JP 2002505881-A 1 26-FEB-2002;  
 DOW AGRSCIENCES LLC  
 COMMENT OS Saccharopolyspora spinosa  
 PN JP 2002505881-A/1  
 PD 26-FEB-2002  
 PF 16-FEB-1999 JP 2000535754  
 PR 09-MAR-1998 US 09/036987  
 PI RICHARD H BALTZ,M CHRISTINE BROUGHTON,KATHRYN P CRAWFORD,PI  
 KRISHNAMURTHY MADDURI,DONALD J MERLO,PATTI J TREADWAY,JAN R  
 TURNER,  
 PI CLIVE WALDRON  
 PC C12N15/09,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P19/62,C12Q1/  
 PC 68//  
 PC C07K14/195,C12N15/00,C12N5/00  
 CC Biosynthetic genes for spinosyn insecticide production FH  
 Key Location/Qualifiers  
 FT source 1. 80161  
 FT Location/Qualifiers  
 /organism='Saccharopolyspora spinosa'.  
 1. 80161  
 /organism="Saccharopolyspora spinosa"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:60894"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2.93e-98 Length: 80161  
 Score: 1444.00 Matches: 275  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 Db: 6 Gaps: 0

US-10-069-353A-8 (1-275) x BD137649 (1-80161)  
 QY 1 ValLeuProGlyValAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20  
 Db 20168 GTGTGGCCAGGTGGGCGACCAACATCGCAGCAGGTGGGCGAGATGATGACTGTGTCACG 20227  
 QY 21 ProLeuLeuAsnSerValAlaGlyGlyProCysAlaAlaIleHisGlyTyrTrpGluAsn 40  
 Db 20228 CCGTTGCTGMACTGGGTGTGGGGGGGGCCCTGGCCATCCACCAACGGCTTACGGGAGAAC 20287  
 QY 41 AspGlyValAlaSerTyrGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60  
 Db 20288 GACGGCGGGGCTCTTGGCCACAGCGCGCCACCGGCTCACCGACCTTGTGCGGAACGG 20347  
 QY 61 ThrValLeuAspGlyValValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80  
 Db 20348 ACCGTGCTGATGGCGGGCTTCGACTGCTCGATGTGGGTGCGGTACCGGACACACGG 20407  
 QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100  
 Db 20408 CTGGCGGTGGCGCGGACCAACCGCATCCAGATCACCGCATCACCGTCCAGCGAGTGCAA 20467  
 QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120  
 Db 20468 GTGGCCATCGCGCTGATTGGCGACCGCGAACCGGACTAACCCACCGGCTGGACTTCTCG 20527  
 QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140  
 Db 20528 TCGTGCATGCCATGTCTCTCTCCCTGACCCGGAATGCTTTGACGCCGCTGGAGCCATG 20587  
 QY 141 GluSerLeuLeuGluMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160  
 Db 20588 CAGTGCCTGTTGGAGATGTCCGAACCGGACCGGTGCATCCGGAAATCTTCGATGACTC 20647  
 QY 161 LysProGlyValIleuGlyValIleThrGluValValIleArgGluAlaGlyGlyMet 180  
 Db 20648 AAACCCGGTGCATCTCTCGGGTCAACCGAGTCTGTCAACGAGACGGGGCGGATG 20707  
 QY 181 ProValSerGlyAspArgTrpProThrglyLeuValIleCysLeuAlaGluGlnLeu 200  
 Db 20708 CCGGTGTCCGGGAGCAGGTGGCGGACCGGCTTCGGATCTGCTGGCTAGCACTTCTG 20767  
 QY 201 GluSerLeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220  
 Db 20768 GAATGCTGCGTGCAGCGGGGTTCAGATCTCTCGATTGGAGAGCGTGTGCTCAGAGACC 20827  
 QY 221 ArgTyrPheMetProGlnPheAlaGluGlnLeuAlaAlaHisGlnHisGlyIleAlaAsp 240  
 Db 20828 CCGTACTTCATGCCGAGTTCCCGAAGAGCTCGCTGGCCACACAGACGGGATCCGGAC 20887  
 QY 241 ArgTyrGlyProAlaValAlaGlyTyrTrpAlaAlaAlaValCysAspTyrGluLysTyrAla 260  
 Db 20888 AGGTACGGGCGCGCTGTCCGCGGTGGCGCGCGCGCTGCTCGATTATGAGAAATATGCC 20947  
 QY 261 HisAspMetGlyTyrAlaAlaIleuThrAlaArgLysProValGly 275  
 Db 20948 CACGACATGGGCTATGCGATCTGACGCGCGGAGACCGGTGCGC 20992  
 RESULT 7  
 LOCUS AR281866 80161 bp DNA linear PAT 10-APR-2003  
 DEFINITION Sequence 1 from patent US 6521406.  
 ACCESSION AR281866  
 VERSION AR281866.1 GI:29717767  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 80161)  
 AUTHORS Baltz,R.H., Broughton,M.C., Crawford,K.P., Madduri,K., Merlo,D.J.,  
 Treadway,P.J., Turner,J.R. and Waldron,C.  
 TITLE SpnG, a gene for spinosyn insecticide biosynthesis  
 JOURNAL Patent: US 6521406-A 1 18-FEB-2003;



FEATURES  
 source  
 Location/Qualifiers  
 1..80161  
 /organism="unknown"  
 /mol\_type="genomic DNA"

ORIGIN

Alignment Scores:  
 Pred. No.: 2,93e-98 Length: 80161  
 Score: 1444.00 Matches: 275  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0

US-10-069-353a-8 (1-275) x AR281866 (1-80161)

QY 1 ValLeuProGlyVgLYAlaProThrsSerGlnGlnValGlyGlnMetYrAspLeuValThr 20  
 Db GTGTTGCCAGGTGGCGGCAACCAATCGACAGCTTGCGCATGTATGACCTGCTCAGC 20227

QY 21 ProLeuLeuAsnSerValAlaGlyProCyAlaIleHisHsGlyYrTPGluAan 40  
 Db CCGTTGCTGAACCTCGGTCCGGCGGCGCCCTGCGCATCCACACGCGTACTGAGAAAC 20287

QY 41 AspGlyArGAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60  
 Db GACGGCGGGCTTCTTGCGACAGCGCGCGACCGGCTCACACCTTGTCCGCGAACGG 20347

QY 61 ThrValLeuAspGlyVgLYAlaArgLeuLeuAspValGlyCyArgLYThrGlyGlnProAla 80  
 Db ACCGTGCTGCATGCGCGGCTTCGATGCTGATGTGGGGTGGCGTACCGGACCAACGCG 20407

QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100  
 Db CTGCGGCTCGCGGCGGCAACCGGATCCAGATCACCGGATCACCGTCACAGGTGCA 20467

QY 101 ValAlaIleAlaAlaAspCyAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120  
 Db GTGGCCATCCCGCTGATTCGCGACCGGCAACCGGATTAAGCCACCGGATGACTTCTCG 20527

QY 121 CyValAlaAspAlaMetSerLeuProYrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140  
 Db TGGGTGCATGCCATGCTCCGCGTACCGGACATCTTTCAGCCCGCTGGGCCATG 20587

QY 141 GlnSerLeuLeuGlnMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160  
 Db CAGTCCGTGTGGAGATGTCGAAACCGGACCGGATCCGGAATCTTCCAGTACTC 20647

QY 161 LysProGlyGlyIleLeuGlyValThrGluValValLysArgGluAlaGlyGlyGlyMet 180  
 Db AAACCCCGTGCATCTCCGCGGTCAACGAGGTCTCAACGAGAGGCGGCGGGGATG 20707

QY 181 ProValSerGlyAspArgTrpProThrGlyLeuArgIleCyAlaLeuAlaGlnLeuLeu 200  
 Db CCGGTGTCGGGGGACAGGTGCGACCGGCTTTCGATTCCTGCGCTGAGCAACTTCTG 20767

QY 201 GluSerLeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220  
 Db GAATCGCTGGGTGAGCGGGGTTCGAGATCTTCATTTGGAGGAGCGTGTCTGAGAAC 20827

QY 221 ArgTrpPheMetCProGlnPheAlaGluGluLeuAlaAlaHisArgHisGlyIleAlaAsp 240  
 Db CGGTACTTCAATGCCGAGTTCGCGAAGAGCTCTCTCGCACCCGAGCAGGAGATCGCGAC 20887

QY 241 ArgTrpGlyProAlaValAlaGlyTrpAlaAlaAlaValCyAspArgGluLysYrAla 260  
 Db AGGTACGGGCGCGCTGTGCGCGGCTGGCGCGCGCGCTGCGATTATGAGAAATATGCC 20947

QY 261 HisAspMetGlyYrAlaIleLeuThrAlaArgLysProValGly 275  
 Db CACAGCATGGGCTATGCGATTCTGACGGCGGAGCGGATCGGC 20992

RESULT 8

AX600587/C  
 LOCUS AX600587 36538 bp DNA linear PAT 14-FEB-2003  
 DEFINITION Sequence 2 from Patent WO02079477.  
 ACCESSION AX600587  
 VERSION AX600587.1 GI:28400303  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Saccharopolyspora sp. NRRL 30141  
 Saccharopolyspora sp. NRRL 30141  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Pseudonocardiales; Pseudonocardiaaceae; Saccharopolyspora.

REFERENCE  
 1 Hahn, D.R., Jackson, J.D., Bullard, B.S., Gustafson, G.D., Waldron, C. and Mitchell, J.C.  
 Biosynthetic genes for butenyl-spirosyn insecticide production  
 Patent: WO 02079477-A 2 10-OCT-2002;  
 Dow Agrosciences LLC (US)

TITLE  
 JOURNAL  
 Location/Qualifiers

FEATURES  
 source  
 1..36538  
 /organism="Saccharopolyspora sp. NRRL 30141"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:221471"  
 /note="LM107129"

ORIGIN

Alignment Scores:  
 Pred. No.: 7,04e-89 Length: 36538  
 Score: 1315.00 Matches: 252  
 Percent Similarity: 94.18% Conservative: 7  
 Best Local Similarity: 91.64% Mismatches: 16  
 Query Match: 91.07% Indels: 0  
 DB: Gaps: 0

US-10-069-353a-8 (1-275) x AX600587 (1-36538)

QY 1 ValLeuProGlyVgLYAlaProThrsSerGlnGlnValGlyGlnMetYrAspLeuValThr 20  
 Db GTGTTGCCAGGTGGCGGCAACCAATCGACAGCTTGCGCATGTATGACCTGCTCAGC 879

QY 21 ProLeuLeuAsnSerValAlaGlyProCyAlaIleHisHsGlyYrTPGluAan 40  
 Db CCGTTGCTGAATTCGGTCCGGCGGCGCCCTGCGCATCCACACGCGTACTGAGAAAC 819

QY 41 AspGlyArGAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60  
 Db GACGGCGGACTTCTTGCGACAGCGCGCGACCGGCTCACCGATTTGCGCGAACGG 759

QY 818 ThrValLeuAspGlyVgLYAlaArgLeuLeuAspValGlyCyArgLYThrGlyGlnProAla 80  
 Db ACCGCGCTGCATGCGGCAATGACTGCTGATGTGGGTGCGGACCGGACCAACGCGG 699

QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100  
 Db CTGCGGCTCGCGGCGGCAACCGGATCCGAGTCAACGAGTCAACGAGTCAACGAGTCA 639

QY 698 ThrValLeuAspGlyVgLYAlaArgLeuLeuAspValGlyCyArgLYThrGlyGlnProAla 120  
 Db GCTGCTGCATGCCCTTATGCGCGCGCGGACCGGATTAAGCCACCAAGTGCATCTCTG 579

QY 101 ValAlaIleAlaAlaAspCyAlaArgGluArgGlyLeuSerHisArgValAspPheSer 140  
 Db GCTGCTGCATGCCCTTATGCGCGCGCGGACCGGATTAAGCCACCAAGTGCATCTCTG 519

QY 121 CyValAlaAspAlaMetSerLeuProYrProAspAsnAlaPheAspAlaAlaTrpAlaMet 160  
 Db TGGGTGCATGCCATGCTCCGCGTACCGGACCGGATCCGGAATCTTCCAGTACTC 459

QY 578 GluSerLeuLeuGlnMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 180  
 Db CAGTCCGTGTGGAGATGTCGAAACCGGATCCGGAATCTTCCAGTACTC 399

QY 161 LysProGlyGlyIleLeuGlyValThrGluValValLysArgGluAlaGlyGlyGlyMet 200  
 Db AAACCCCGTGCATCTCCGCGGTCAACGAGGTCTCAACGAGAGGCGGCGGGGATG 339

QY 181 ProValSerGlyAspArgTrpProThrGlyLeuArgIleCyAlaLeuAlaGlnLeuLeu 200  
 Db CCGGTGTCGGGGGACAGGTGCGACCGGCTTTCGATTCCTGCGCTGAGCAACTTCTG 339





```
/db xref="GI:29604394"
/translation="MGTEITSPAOVGLRESKKEHTROLISDHATRLFLAAGEEOTTI
ARIAAABMAKKTVTNYFARKEDIALDHOEFPLASIANVADRRPGESLALAFPH
DLATONPYAFSGSFAFAMVADSPITLTCGFGHDOREHRLAQALADDTDTTDPDIT
VHIAAGLGTFRVLFQRIQDLTLAGHPNDRIAHITTAETAQAFDLLEPAHAHVATA"
4587. .5627
/feature="SAV755"
/codon_start=1
/translation="putative dehydrogenase"
/protein_id="BAC68465.1"
/db xref="GI:29604395"
/translation="MTSPKRTTTTSPVACPATGSPSQATSPAPYPAVTRACSTWT
CSRSDIRPAPADVPSPQADFEADFPALVARTLSDLTPGERHYLRYSVLDSFDP
LYTRAGTADAPALRLTERPVRNPGTCGGAAPAPADPDMAVSGRPARSDGP
APTSSRPASPAAPGPHOAGSAVGMTEVTPKEMAYAVASLIDVHTHGRALPHR
AARGSNVSGAGLGGDYGMAAYNAKGAVNTTRAMALDHAEVRYNAGHPGPTTHSG
ITAGATPDTPVVRSFAERIPLGRVAAPDEIASVAVFALGPDALYDRSRNPGREPHRC
YAAAPHDVNA5"
complement(6101. .8071)
/feature="SAV756"
/codon_start=1
/translation="hypothetical protein"
/protein_id="BAC68466.1"
/db xref="GI:29604396"
/translation="MERSKDRRPSVLODRARYMPDRTLARSNTGLMGLVITICLAVV
PSALLVMTDPSGSPRSLGTRILAAVWSSAFTRLCTVGTCTPRMLISLIGLVALCY
STLVITITGLAERLAEISRGSTVLEBQHAIVLMSDOVSTVSELYAAGSITPRA
VVLAEKDTTEMERLAAHVGPAGTRILCRSGPASDPEVLAVSPQATVTVLPBG
EPTADAEVRLVLLRAVLEGSTGPVLAALADRYAPALAPALGAPGTVLETDTVT
ALIAQCVGRPGSLVRLDLDFADDEFLEATAPFGHFGALALGHTASCVVLT
AEGRTLPATLVAPCSRILVLTREDDSARPEDCRHLVESSALAMQPEDPAHL
ILGNNRAPLIVNOLRTPARPSVLDVTDRAVPGRPRESSGAARLAVRQSDSL
RPTTLGLDIAPYDGLIVGPRPGDGPREDRTIYTLALRLERGRERIVTEL
TDRRRLPAPVPGSDVIVSGELTGLLQAQIAQNHLAPFDELFSADGSTICLQAG
HVRGTCEETFAVVAADRGECALGYRRHDIRAI PDHGIRLNPDSESRREMSAD
QVVVAITEPTEPAAGQDTEBHOEPADTRA"
5086. .10072
/feature="SAV757"
/codon_start=1
/translation="hypothetical protein"
/protein_id="BAC68467.1"
/db xref="GI:29604397"
/translation="MGRKKGIGIALVTVSVMVGASACGGSGDANGDDAKASGPG
ASAPSPKKPAPGAPWLLETITPQCTTGVAMPISVVTNNPRAARAARAEKHMVSA
DVRADVSVTGHMKVTRNGKVTRLISNAGSAQYPTNGTMAVIDKQEKHMTSCVY
GISCEKSGPNVYDITLPMDVHLTQGTIVHVSIGDPTGSGSGARSGHCVHLMSMDAK
WFYGVQKQDPTLTIGSPRAKAPADNGVADFNLMGWDQMLAGSASGEKTTATY"
complement(10889. .11494)
/feature="SAV758"
/codon_start=1
/translation="putative acetyltransferase"
/protein_id="BAC68468.1"
/db xref="GI:29604398"
/translation="MORAVPRAIMAVMARPOLIECGDENVLRSMRGODPAPAFYLIE
ESDLRLRMEPVAHSESESTRDLAKESKRANDVNYIALKQDITLIGTCOTYGA
EPQGRIGIWMHPATGGIATRAAAMTEMTLTPDEVLEVADLNTSSAIPRR
LGFEVREQVPPAPASGSGIDVVMRLNRPPTVPDGSRRP"
11580. .12161
/feature="SAV759"
/codon_start=1
/translation="hypothetical protein"
/protein_id="BAC68469.1"
/db xref="GI:29604399"
/translation="MRSRAPTRCMMWPSSTRPSDATALCMRWMTGCTGTGPRPRDI
SAACFRAPVALLTOARPGKILLITSGMUYGARVLLTVADQGTAAQGLRGPVSR
MSAPHAAPTYGNPAHNAATGLPEKARTTRHSPERTTITDDEALHLRFLAQASVSDAQM
TTVDVHVLVDVVTGSRDPHPRGQLGCTRGSR"
```

```
CDS
complement(12097. .12768)
/feature="SAV760"
/codon_start=1
/translation="putative multicopper oxidase"
/protein_id="BAC68470.1"
/db xref="GI:29604400"
/translation="MQAFSPHFAEDYATLEPSRAKPNENAIYGTNNHRSOMLYHD
HAMGSLANVYAGLGLVLRDPPDERGLPRGAFEVLLIRDPFRNDGSLATMTA
RGEEDTPVFNKAPFLAVERRRLRLTINSNEPFRRLRDVPADVLLQQLPFWLI
GTDGGEFRAPLKMLDFLISSAERYDLIDVSGMPRTSRIRCHSFTGPLAYSPTVLARG
GRG9Q"
complement(12621. .12968)

Alignment Scores:
Pred. No.: 8, 77e-22 Length: 302675
Score: 441.50 Matches: 111
Percent Similarity: 56.32% Conservative: 45
Best Local Similarity: 40.07% Mismatches: 107
Query Match: 30.57% Indels: 15
DB: 1 Gaps: 6

US-10-069-353A-8 (1-275) x AP005024 (1-302675)

QY 7 ProThrSerGlnGlnValGlyGlnMetYrAspLeuValThrProLeuLeuAsnSerVal 26
Db 237453 CCTCTTCCCTCGAAGTGGGGGAGACTACGACCGTTGACCGACCTCATGATTCGTGCG
QY 27 AlaGlyGlyProCysAlaIleHisGlyTyrTrpGluAsnAspGlyArgAlaSerTrp 46
Db 237393 CTGGGTGG-----ACACCCACCCTCGATACGTGGCCGACGCGGGGAGCGGAGTTCA
QY 47 Gln---GlnAlaIleAspAspLeuThrAspLeuValAlaGluArgThrValLeuAspGly 65
Db 237339 CCGGCAAGCGCCGCGACCGGCTACCGATCTCTCATCGGCAAGCTGACGACATCACG
QY 66 GlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAlaLeuArgValAlaArg 85
Db 237279 GGC CGCGCGGTCTGGACGTCGTGGCGGTCCGGAAGACCGGGGCGTGGCGCCCTG
QY 86 AspAsnAlaIleGlnIleThrGlyTleThrValSerGlnValGlnValAlaIleAlaAla 105
Db 237219 AGCGGCGCGCGTGAATGTCGTGGCGGTGACGGTGAAGGTTCAGGTGCGCGGCGAC
QY 106 AspCysAlaArgGlyLeuArgGlyLeuSerHisArgValAspPheSerCysValAspAlaMet 125
Db 237159 GCTCTCGGAGACGATCGCACGTGGCGGACCGGGTCTGTTCACCGGTCGCGGCGGATG
QY 126 SerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMetGlnSerLeuLeuGlu 145
Db 237099 GAACGTGCGGTTTCCGACGCGGTCTTGACGCGGGCGGTGGCCCTGGAAGTGTCTCTGAC
QY 146 MetSerGluProAspArgAlaIleArgGluIleLeuArgValIleLeuLysProGlyGlyTle 165
Db 237039 ATGCCAGACCCCGCACAGGTATCCGGGAGATCCCGGGGTGCTCCGCGCGGCGGCGG
QY 166 LeuGlyValThrGluValIleValArg-GluAlaGlyGlyGlyMetProValSerGlyAs 185
Db 236979 CTGGCGCTGACGACGTCGACCTGCGGCTTCGGGGGAGACCGGCAAGAACCGGGGAG
QY 185 PArgTrpProThrGlyLeuArgIle-----CysLeuAlaGluGlnLeuLeuGluSe 202
Db 236919 TGCAAGTCCAGTGTCTGCGCGGTCCCGGCTCTGTGACATCGACGAGTACCGCGAATG
QY 202 rLeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThrArgTyr 222
Db 236859 -ATGCCGACCCCGGCGGTGGAATCGATGACCTGACCGACATCGCGCATGCGT-----
QY 222 rPheMetProGlnIlePheAla-----GluGluLeuAlaAlaHisGlnHisGlyTleAl 239
Db 236805 -GTGGGCGGCTCTTTCGCGCGGTGCGGTGACACGTGAACGACGACCTTCGACGAGTACG 236747
QY 239 aAspArgTyrGlyProAlaValAlaGlyTrp-----AlaAlaAlaValCysAspTyrGly 257
```



DB:	6	Gap:	6
US-10-069-353A-8 (1-275) x E07846 (1-1410)			
QY	7	ProthrSerGlnGlnValGlyGlnMetTyrAspLeuValThr-ProLeuLeuAsnSerVal	26
Db	841	CCCTCTTCCCTCGAAGTGGGGGACCTACACGACCGTTTACCGACCTCATGAATCGTGG	782
QY	27	AlaGlyGlyProCysAlaAlaIleHisGlyTyrTPGluAsnAspGlyValArgIaSerTrp	46
Db	781	CTGGGTGGG-----AACACCCACCTCGGATCTCGCGCAGACCGGGGGACGCGAGTTCA	728
QY	47	Gln---GlnAlaAlaAspArgLeuThrAspLeuValAlaGluArgThrValLeuAspGly	65
Db	727	CCGGGCAAGCGCCGACCGGCTCACCATCTCTTCATCGGACAGCTGAGGACATCAG	668
QY	66	GlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAlaLeuArgValAlaArg	85
Db	667	GGCCCCCGGGTCTCGAAGCTCGCGCTCGGGTTCCGGAAGAAGCCGGCGGTGGCGTCCGCTG	608
QY	86	AspAsnAlaIleGlnIleThrGlyLysIleThrValSerGlnValGlnValAlaIleAla	105
Db	607	AGCGGCGCCGTGATGTCGTGGCGTGAAGCGAGACGGTTCAGTCCGGCTGGCGAGCC	548
QY	106	AspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSerCysValAspAlaMet	125
Db	547	GCTCTCGGGAAGCACTCGACGCTGGCGAGACCGGGTCTGTTTACCCGTCCGACCGGATG	488
QY	126	SerLeuProTyrProAspAsnAlaPheAspAlaAlaATPAlaMetGlnSerLeuGlu	145
Db	487	GAACCTCCGGTTCCTCCGACGGGTCTTCGACGGCGGGCGCCCTCGAAGTCTCTCGCAC	428
QY	146	MetSerGluProAspArgAlaIleArgGluIleLeuArgValLeuAspProGlyIle	165
Db	427	ATGCCACAGCCCGGACAGGTATCCGGGAGATCGCCGGGTCTCGCCGCGGCGG	368
QY	166	LeuGlyValIleThrGluValValIleAspArgGluAlaGlyGlyMetProValSerGlyAsp	185
Db	367	CTGGCGGTGACGAGCATGCTACGCGGCTT-TCGGGCGGACCGGACATGAAGCCGGGGAGT	309
QY	186	ArgTrpProThrGlyLeuArgIle-----CysLeuAlaGluGlnLeuGlnSer	202
Db	308	GCACGCTCCAAAGTGTCTCGGGGATCCCGGCGCCCTGTGTCATCAACAGATACCGGGAATG-	250
QY	203	LeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThrArgTyr	222
Db	249	ATCGCGACGCGCGGCTGGAATCTGATGATGACGACCATCGGGGATGAGTCTC-----	196
QY	223	PheMetProGlnPheAla-----GluGluLeuAlaAlaHisGlnHisGlyIleAla	239
Db	195	GTCGGCGCCCTCTTTGGCCGCGCTGGCTGACACGTAACGAGACACTCGACGATGACCG	136
QY	240	AspArgTyrGlyProAlaValAlaGlyTyr-----AlaAlaAlaValCysAspTyrGlu	257
Db	135	GCGGCTTTCGGATGCGGGCGGACGAGATCGGAGACGCTTGTCACAGTGCACGACGCTC	76
QY	258	LysTyrAlaHisAspMetGlyTyrAlaIleLeuThrAlaArgLysPro	273
Db	75	CCCTGACGCGGACATCGGCTATGTCGTGCTGACCGGCCGCGCGCGC	28
RESULT 13			
AX598613			
LOCUS	AX598613	846 bp	DNA linear
DEFINITION	Sequence 21 from Patent WO02088176.		PAT 14-FEB-2003
ACCESSION	AX598613		
VERSION	AX598613.1		
KEYWORDS	GI:28398761		
SOURCE	Streptomyces platenis		
ORGANISM	Streptomyces platenis		
REFERENCE	Bacteria; Actinobacteria; Actinobacteridae; Actinometales;		
AUTHORS	Streptomycineae; Streptomycetaceae; Streptomyces.		
	1 Farnet, C.M., Zazopoulos, E., Staiffe, A. and Yang, X.		

TITLE		Genes and proteins for the biosynthesis of polyketides	
JOURNAL		Patent: WO 0208176-A 21 07-NOV-2002;	
FEATURES		Ecopia Biosciences Inc. (CA)	
source		Location/Qualifiers	
		1..846	
		/organism="Streptomyces platensis"	
		/mol_type="unassigned DNA"	
		/db_xref="taxon:58346"	
		/note="subsp. rosaceus"	
ORIGIN			
Alignment Scores:			
Pred. No.:	2,21e-24	length:	846
Score:	438.50	Matches:	106
Percent Similarity:	56.108	Conservative:	55
Best Local Similarity:	36.93%	Mismatches:	95
Query Match:	30.37%	Indels:	31
DB:	6	Gaps:	9
US-10-069-353A-8 (1-275) x AX598613 (1-846)			
QY	2	LeuPProGlyValAlaProThrSerGlnGlnValAlaGlyMetTyAspLeuValThrPro	21
DB	34	GTCCCGGTCCTCCGCCCGCCCTCCCGAAGAGTGTGGACACTTCACACCGCCACCGCA	93
QY	22	LeuLeuAenSerValAlaGlyGlyProCysAlaIleHisGlyTyrThrGlu---Asn	40
DB	94	CTGGACACCGAAGCGCGCGCGGC-----AGCCTTCACCTGGCTACTGGACGTGCAC	147
QY	41	AspGlyArgAlaSerTyrGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg	60
DB	148	GACAACGACACCCCGCTGTGGAAAGGGCGACCGGCTCACGACACGATGACCGACCGC	207
QY	61	ThrValLeuAspGlyGlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla	80
DB	208	CTGCCGATGACACGAGGACGAGCGGGCTCTGCAGCTGGCTGGAGACTCGACACGCGGCC	267
QY	81	LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln	1001
DB	266	ATGGCGGATGGCCCGGGGACCGGCGCCCATGTACAGGGCATGGCATGACGAAAGACAG	327
QY	101	ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer	120
DB	328	ATGCGCCCGGACCGCCCTCCGCCAGAGGCGCGGCTGAGGACGACCGCGTGAAGTTCCG	387
QY	121	CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet	140
DB	368	CACGCGGACGATGAGACTGCTCCCTTCCCGACGACTCTTGACGCCGCCCATGCGCATC	447
QY	141	GlnSerLeuLeuGlnMetSerGlnProAspArgAlaIleArgGluIleLeuArgValLeu	160
DB	448	GAGTGCATCTTCACATCGCCGACCGCGGACGGGATCTGCGGAGATCCGCGCGTACAG	507
QY	161	LeuPProGlyGlyIleLeuGlyValThrGluValValLeuArgGluAlaGlyGlyMet	180
DB	508	CGCCCGGGGCGCGCTGCTCTCACCGACTTCTTGACGCGGCGC-----	552
QY	181	ProValSerGlyAspArgTyrProThrGly-----LeuArg	192
DB	553	CCCGTCCCGCGGAGAGAGCGCCGGGGTGGACCGGCTCTCCGCACTTCATCATGACG	612
QY	193	IleCysLeuAlaGlyGlnLeuLeuGlnSerLeuAlaGlyAlaGlyPhe-----Glu	209
DB	613	CTGGCCCGCGCCGAGAGACTACGTGCCATGTGCGCGAGCGAGGCGCTGCGCTTGTGAG	672
QY	210	IleLeuAspTyrGluAspValSerSerArgThrArgTyr---PheMetProGlnPheAla	228
DB	673	CTCTCTGACATCACCGAGAGAGCGTGGCTCAGACCTTGACAGCAATGAGCCAGGGCTCC	732
QY	229	GluGluLeu---AlaAlaHisGlnHisGlyIleAlaAspArgTyrGlyProAlaValAla	247
DB	733	CAGGAGATGACAGACCGTCTTGACGACGAGGACGAGAAAGTTTCAGGCC-----	783

Oy		248	GlyTrpAlaIleAlaValCysAspTryxGluYtyrValAlaHisAspMetGlyTyrAlaLeu	267
Dd		784	-----GCCCTCATGATCGACGTCCAGCAATTCTC-----GGCTTCGTTCTG	822
Oy		268	LeuThrAlaArgLysProVal	274
Dd		823	CTGACC GCCCAAAGCCCCCTC	843
RESULT 14				
Locus		AX598593		
DEFINITION		Sequence 1 from Patent WO02088176.		
LOCUS		AX598593	52101 bp	DNA linear PAT 14-FEB-2003
VERSION		AX598593.1	GI:28398751	
KEYWORDS				
SOURCE		Streptomyces platensis		
ORGANISM		Streptomyces platensis		
REFERENCE		Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. Farnet,C.M., Zazopoulos,E., Staffa,A. and Yang,X. Genes and proteins for the biosynthesis of polyketides Patent: WO 02088176-A 1 07-Nov-2002;		
AUTHORS		Ecopia Biosciences Inc. (CA)		
TITLE		Location/Qualifiers		
JOURNAL		1..52101 /organism="Streptomyces platensis"		
FEATURES		/mol_type="unassigned DNA"		
SOURCE		/db_xref="taxon:58346"		
ORIGIN		/note="Subsp. rosaceus"		
Alignment Scores:				
Pred. No.:	2,11e-22	Length:	52101	
Score:	438.50	Matches:	106	
Percent Similarity:	56.10%	Conservative:	55	
Best Local Similarity:	36.93%	Mismatches:	95	
Query Match:	30.37%	Indels:	31	
DB:	6	Gaps:	9	
US-10-069-353A-8 (1-275) x AX598593 (1-52101)				
Oy		2	LeuProGlyAlaIleAProThrsSerGlnGlnValGlyGlmMetYrrAspLeuValThrPro	21
Dd		51267	GTCCCGCTCTCCGCCGCCCTCCCGAGAAGAGTGTCGAACCTTCACACCGCTCACCGCA	51326
Oy		22	LeuLeuAenSerValAlaGlyGlyProCysAlalIleHisIsIGlyTYrTPGlu--Aen	40
Dd		51337	CTGCAGACCCGAAGCGCGCGCGGC-----AGCTCCACCTCGGACTCTGCGACGTCCAC	51380
Oy		41	AspGlyArgAlaSerITripGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg	60
Dd		51381	GACAACGACACCCCGCTGTGGAAAGCGCGGACCGGCTCACCGACACGATGACCGACCGC	51440
Oy		61	ThrValLeuAspGlyGlyValArgLeuLeuAspValGlyCyseGlyThyrGlyInProAla	80
Dd		51441	CTGCCGATCGACCAAGGGAACACGGGTCTCTCGACCTCGGCTGCGAGTCCGCGACGCGGC	51500
Oy		81	LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln	100
Dd		51501	ATGCCGATGCGCCCGCGCAGACCGGCGCCCATCTCACGGGCAATTCGGATCGACAGACAGCAG	51560
Oy		101	ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer	120
Dd		51561	ATCCGCCGCGCACCGCCCTCGCGAAGGAGCGCGGCTGAGCGACCGCGTAGATTCCG	51620
Oy		121	CysValAspAlaMetSerLeuProTyrrProAspAsnAlaPheAspAlaAlaTrrPalamet	140
Dd		51621	CACGCCGACGCATGAACTGCCCTTCCCACGACATCTTCGACGCGCGCATCGGCATC	51680
Oy		141	GlnSerLeuLeuGlmMetSerGlnLProAspArgAlaIleArgGluIleLeuArgValLeu	160
Dd		51681	GAGTGTGATCTTTCACATCCCGACCGCGGAAGGGATCTCTCGCGAGATCCCGCGATCTG	51744

Oy		161	LysProGlyIcylIleLeuGlyValThrGluValValLysArgGluAlaGlyIcylGlyMet	180
Db		51741	CGCCCCGGCGCGCCCTGTCTTCACCGACTTCTTCAGCGCGGC-----	51785
Oy		181	ProValSerGIcYAsparGTrgPrroThrgLy-----LeuArg	192
Db		51786	CCCGTCCCCCGGAGAAGCAGCCCGCGGGAACCGGCTCTCCCGCACTTCATCATAGAG	51845
Oy		193	IleCysLeuAlaGlucIleuLeuGluSerLeuAlaAlaGlyPhe-----Glu	209
Db		51846	CTGGCCCGCGGCCAGAGACTAAGCGCCCATGTCCGCGAGCGAGCGCTGCGCTTGCTGAG	51905
Oy		210	IleLeuAsprTrgGluAsprValSerSerArgTrnArgTrg----PheMetProGlnPheala	228
Db		51906	CTCTTGACATCAACCAGCAGACCGCGCTGCTCAGACTTTCAGACAGATGAGCCAGGCTCC	51965
Oy		229	GluGluLeu---AlaAlaHisGlnHisGlyIleAlaAsparGTrgIcylProAlaVala	247
Db		51966	CAGAGAGTGCAGACCGCTTCGACGACGAGCGAGAAAGTTTCAGCCCC-----	52016
Oy		248	GlyTrpAlaAlaAlaValCysAsprTrgGluIcylTrpAlaHisAspmetGlyTralaile	267
Db		52017	-----GCTTCAGATGATGACGTGCAGCAATTG-----GCTTCGTTCTG	52055
Oy		268	LeuThrAlaArgLysProval	274
Db		52056	CTGACCGCCCAAAAGCCCTC	52076
RESULT 15				
LOCUS	AX089434		900 bp	DNA linear PAT 21-MAR-2001
DEFINITION	Sequence 19 from Patent WO0116303.			
ACCESSION	AX089434			
VERSION	AX089434.1		GI:13443695	
KEYWORDS	Saccharopolyspora spinosa			
ORGANISM	Saccharopolyspora spinosa			
SOURCE	Bacteria; Actinobacteriae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.			
REFERENCE	Eberz,G., Moehle,V., Froede,R., Valten,R. and Sales,J.A. Nucleic acids which code for the enzyme activities of the spinosyn biosynthesis Patent: WO 0116303-A 19 08-MAR-2001;			
AUTHORS	Patient: MO 0116303-A 19 08-MAR-2001;			
TITLE	BAAYER AG (DE)			
JOURNAL				
FEATURES				
source	location/Qualifiers			
	1..900			
	/organism="Saccharopolyspora spinosa"			
	/mol_type="unassigned DNA"			
	/db_xref="taxon:60894"			
CDS	1..900			
	/note=unnamed protein product; ORF7; O-Methyltransferase"			
	/codon_start=1			
	/transl_table=11			
	/protein_id="CACJ4880.1"			
	/db_xref="GI:13443696"			
	/translation="MNSRGANGAPPTADQVESIFDALAHGREPLHHGYWAGYREDAG ATPSDADQDTLDLFIDKALIRPGAILFDIGSCNGOPVVAACASGVRTYGIVLNQH IAARTFARNETGLAGLEPDVLVGAOLPYDGFGFOVAMAMQSVOVLDOAAITREYOR ILBEGGFVUGDITTRVRLEBEYAATWTGTTATLTNSFTALVSSEAGFEILEVDLTAA TRCVMSYVELRKIDELAGVEPALAVGTIQRYLGDIAAKHGPFAQLIAVAABERK HPDYARNEBSMGFWLLQARRKS"			
ALIGNMENT SCORES:				
Pred. No.:	3,99e-24		Length:	900
Score:	435.50		Matches:	108
Percent Similarity:	51.67%		Conservative:	47
Best Local Similarity:	36.00%		Mismatches:	96
Query Match:	30.16%		Indels:	49
DB:	6		Gaps:	11

US-10-069-353A-8 (1-275) x AX089434 (1-900)

```
OY 7 ProthSerGlnGlnValGlyGlnMetTyrAspLeuValThrProLeuLeuAsnSerVal 26
    |||||::: |||||::: |||||:::
Db 34 CCGACCGCGGATCAGGTGAGTCCATCTTCGATCGCTTG-----GGG 75
OY 27 AlaGlyGlyProCysAlaIleHisGlyTyrTrp-----GluAsnAsp 41
    |||||::: |||||::: |||||::: |||||:::
Db 76 CACGGGCGTCCC-----CTGCACACCGGTTACTGGGCGGCGGATCGGAGATGCC 129
OY 42 GlyArgAlaSerTyrGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArgThr 61
    |||||::: |||||::: |||||::: |||||::: |||||:::
Db 130 GGTGCCACACCGGTGCGGATGCTGCCGACCAACTGACCGACCTGTCATCGACAAGGCC 189
OY 62 ValLeuAspGlyGlyValArgLeuAspValGlyCysGlyThrGlyGlnProAlaLeu 81
    |||||::: |||||::: |||||::: |||||::: |||||:::
Db 190 GGGCTTCGTCGCGGAGCGGACCTGTTCCGACTGGGCTGGGCAATGGGCAGCCGTAATC 249
OY 82 ArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGlnVal 101
    |||||::: |||||::: |||||::: |||||::: |||||:::
Db 250 CGTGCGGCAATGCGCGCAGCGGCTTCGACGACCGGATCACCGGATCACCGCAGCATCTC 309
OY 102 AlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspSerCys 121
    |||||::: |||||::: |||||::: |||||::: |||||:::
Db 310 GCGCGCGCTCACAGGCTCGCCAAAGACCGGACTGGCCGCGCATCTTGAGTTCAATCTA 369
OY 122 ValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTyrAlaMetGln 141
    |||||::: |||||::: |||||::: |||||::: |||||:::
Db 370 GTCGACGGGCGCCAGTCCCTACCGGACGGTTCTTTCAGCGCGCATGGCGATGCCAG 429
OY 142 SerLeuLeuGluMetSerGlnProAspArgAlaIleArgGluIleLeuArgValLeuIys 161
    |||||::: |||||::: |||||::: |||||::: |||||:::
Db 430 TCCGTCGTGCAGATCGTGACAGCGCGCGGATCCGCGAGGTCCACCGAATCTTGGA 489
OY 162 ProGlyGly-----IleLeuGly-----ValThrGluValValIysArgGluAlaGly 177
    |||||::: |||||::: |||||::: |||||::: |||||:::
Db 490 CCGGGGCGCGGTTGCTCTCGAGACATCATCACTCGGGTTCGACTCCGGAAGAGTAC 549
OY 178 GlyGlyMetProValSerGlyAspArgTyrProThrGlyLeuArgIleCysLeuAlaGlu 197
    |||||::: |||||::: |||||::: |||||::: |||||:::
Db 550 GCGGGGGTT-----TGG-----ACGGGACGACCGCCCATACCTTGAAAC 588
OY 198 GlnLeuLeuGlnSerLeuArgAlaAlaGlyPheGluIleLeuAspTyrGluAspValSer 217
    |||||::: |||||::: |||||::: |||||::: |||||:::
Db 589 AGCTTCACGCGCGTGTGTCAGCGAAGCGGGGTTGAGATTCTCGAAGTCAACGACTCAG 648
OY 218 SerArgThrArgTyrPheMetProGlnPheAlaGluGluLeu----- 231
    |||||::: |||||::: |||||::: |||||::: |||||:::
Db 649 GCACAGACCGAGTGCATGGTCTCTGTGTACGTGCAGAGTTGCTCCGGAACCTGATGAG 708
OY 232 -----AlaAlaHisGlnHis-----GlyIle 238
    |||||::: |||||::: |||||::: |||||::: |||||:::
Db 709 CTCGCCGCGGTGAGCGCTGCGGCTGCGGCACTACAGCAAGGCTACTGGGAGACATC 768
OY 239 AlaAspArgTyrGlyProAlaValAlaGlyTyrAlaAlaAlaValCysAspTyrGluIys 258
    |||||::: |||||::: |||||::: |||||::: |||||:::
Db 769 GCGGCGAAGCACGACGCGGACACAGTATCGCGCGGTGCGGAAATACCGGAAA 828
OY 259 -----TyrAlaHisAsp-----MetGlyTyrAlaAlaLeuThrAlaArgIys 272
    |||||::: |||||::: |||||::: |||||::: |||||:::
Db 829 CATCGGATTACCGCAAAACAGAAAGAAAGAGTGGTTTCATGCTCTCGAGGCTCGAAG 888
```

Search completed: August 29, 2005, 21:22:47  
Job time : 4877 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 29, 2005, 17:33:28 ; Search time 612 Seconds  
(without alignment)  
2660.014 Million cell updates/sec

Title: US-10-069-353A-8

Perfect score: 1444

Sequence: 1 VLPGAPTSQQVGMVDTLT.....YKVAHDMGYALITARKPVG 275

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlh  
-O/cgn2.1/USPTO.spool/US1006353/runat.26082005.172053.14656/app.query.fasta\_1.455  
-DB=N Geneseq.16Dec04 -OPMT=fastap -SUFFX=mg -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNIT=bits -START=1 -END=1 -MATRIX=blonum62 -TRANS=human0.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pc -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MAXLEN=200000000  
-USER=US1006353 @CGN 1.1 470 @runat.26082005.172053.14656 -NCPY=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq.16Dec04.\*

1: geneseqn1980s.\*  
2: geneseqn1990s.\*  
3: geneseqn2000s.\*  
4: geneseqn2001a.\*  
5: geneseqn2001b.\*  
6: geneseqn2002a.\*  
7: geneseqn2002b.\*  
8: geneseqn2003a.\*  
9: geneseqn2003b.\*  
10: geneseqn2003c.\*  
11: geneseqn2003d.\*  
12: geneseqn2004a.\*  
13: geneseqn2004b.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1444	100.0	828	4	AAf88318
2	1444	100.0	45624	4	AAf88315 S. spinos
3	1444	100.0	50000	4	AAf88312 S. spinos
4	1444	100.0	80161	2	AAZ21501
5	1315	91.1	36538	10	ABV75558

6	439.5	30.4	36538	10	ABV75558	ABV75558 Saccharop
7	439	30.4	1410	2	AAQ80524	AAQ80524 Oxidoredu
8	438.5	30.4	846	10	AAZ54227	AAZ54227 Streptomy
9	438.5	30.4	52101	10	AAZ54217	AAZ54217 Streptomy
10	438.5	30.2	900	4	AAf88324	AAf88324 S. spinos
11	435.5	30.2	45624	4	AAf88315	AAf88315 S. spinos
12	435.5	30.2	50000	4	AAf88312	AAf88312 S. spinos
13	435.5	30.2	80161	2	AAZ21501	AAZ21501 DNA fragm
14	379	26.2	6085	3	AAf70153	AAf70153 S. longisp
15	376.5	26.1	852	2	AAZ55802	AAZ55802 S. lavend
16	376.5	26.1	852	10	AAZ10221	AAZ10221 S. lavend
17	376.5	26.1	53500	3	AAZ55842	AAZ55842 Complete
18	376.5	26.1	53500	10	AAZ10261	AAZ10261 S. lavend
19	375.5	26.0	25681	4	AAZ53165	AAZ53165 Saccharo
20	361.5	25.0	103599	4	ABX04971	ABX04971 S. cinam
21	345	23.9	828	3	AAZ55801	AAZ55801 S. lavend
22	345	23.9	828	10	AAZ10220	AAZ10220 S. lavend
23	343.5	23.8	12381	3	AAZ58381	AAZ58381 Streptomy
24	320.5	22.2	2122	2	AAf70152	AAf70152 S. longisp
25	268	18.6	52659	9	AAf60188	AAf60188 Paederus
26	250	17.3	843	8	ABZ80088	ABZ80088 Anabaena
27	250	17.3	843	10	AAZ19104	AAZ19104 Anabaena
28	247.5	17.3	858	13	AAZ44187	AAZ44187 Bacterial
29	247.5	17.1	7205	13	AAZ84263	AAZ84263 Aspergill
30	246.5	17.1	843	8	ABZ80087	ABZ80087 N. puncti
31	246.5	17.1	843	10	AAZ19103	AAZ19103 Nosoc pu
32	242.5	16.8	1444	6	AAZ23672	AAZ23672 Nicotiana
33	241.5	16.7	1444	6	ABZ53042	ABZ53042 Tobacco b
34	240	16.6	930	8	ABZ80091	ABZ80091 Synechoco
35	238.5	16.5	1383	3	AAZ92614	AAZ92614 Maize C-2
36	232	16.1	1044	8	ABZ80079	ABZ80079 Brassica
37	232	16.1	2590	10	AAZ78872	AAZ78872 Rapeseed
38	229.5	15.9	691	10	AAZ54920	AAZ54920 Plant DNA
39	229.5	15.9	691	10	AAZ58920	AAZ58920 Plant DNA
40	229.5	15.9	1038	8	ABZ80075	ABZ80075 G. hirsut
41	229.5	15.9	1038	10	AAZ19093	AAZ19093 Gossypium
42	229.5	15.9	1038	12	AAZ095177	AAZ095177 Cotton ga
43	229.5	15.9	7713	12	AAZ095182	AAZ095182 PMON75571
44	229.5	15.9	8179	12	AAZ095181	AAZ095181 PMON75565
45	228	15.8	1044	8	ABZ80080	ABZ80080 Brassica

## ALIGNMENTS

RESULT 1  
ID AAF88318 standard; DNA; 828 BP.  
XX  
AC AAF88318;  
XX  
DT 28-AUG-2001 (first entry)  
XX  
DE S. spinosa DNA fragment encoding ORF1, SEQ ID 7.  
XX  
XX Forosamine; trimethylurhamnose; polyketide synthase; biosynthesis;  
XX  
XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;  
XX  
XX macrolide; insecticide; O-methyltransferase; ds.  
XX  
XX Saccharopolyspora spinosa.  
XX  
XX DE19957268-A1.  
XX  
XX PD 08-MAR-2001.  
XX  
XX PF 29-NOV-1999; 99DB-01057268.  
XX  
XX PR 27-AUG-1999; 99DB-01040596.  
XX  
XX PA (FARB ) BAYER AG.  
XX  
XX PI Eberz G, Moehle V, Froede R, Velten R, Sales JA;  
XX  
XX WPI; 2001-267102/28.

DR P-PSDB; AAB70948.

XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for  
PT recombinant production of insecticidal spinosyns and their derivatives.

XX Claim 7a; Page 102-104; 354pp; German.

XX This invention describes a novel method nucleic acid (I) and its encoded  
CC polypeptide (II) containing at least one region that encodes an enzymatic  
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to  
CC identify, inactive or modulate genes involved in the biosynthesis of  
CC (ii); (ii) to generate a library of polypeptide synthases; (iii) for  
CC adding forosamine or trimethylrhinamose to a spinosyn or polypeptide  
CC aglycone; and (iv) for recombinant production of the corresponding  
CC enzymes, which are used for production of (ii), their precursors or  
CC derivatives, including production of transgenic plants that express (ii)  
CC and thus have increased resistance to insects. (I) are also useful as  
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are  
CC macroolides with insecticidal, but not antibacterial, activity, and can  
CC also be used to raise specific antibodies, useful for identifying  
CC expression clones in a gene bank. Cells transformed with (I) may produce  
CC (II) at significantly increased levels or produce new derivatives of  
CC (II). This sequence, ORF 1, encodes an S. spinosa O-methyltransferase  
CC protein

XX SQ Sequence 828 BP; 140 A; 262 C; 285 G; 141 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8 36e-133 Length: 828  
Score: 1444.00 Matches: 275  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-069-353a-8 (1-275) x AAF88318 (1-828)

QY 1 ValLeuProGlyValAlaProThrsErgInGlnValGlyGlnMetTyrAspLeuValThr 20  
DB 1 GTGTTGCCAGTGGCGCACCAATCGACGAGTTGGGCAATGATGACCTGTGTCAGC 60  
QY 21 ProLeuLeuAenSerValAlaGlyProCySaAlaIleHisGlyTyrTrpGluAsn 40  
DB 61 CGTTGCTGAATCGGTGGCGGCGCCCTCGCATCCACACGGCTACTCGGAGAAC 120  
QY 41 AspGlyArgAlaSerTrpGlnGlnAlaAspArgLeuThrAspLeuValAlaGluArg 60  
DB 121 GACGGGCGGGCTTCTGCGACAGAGCGCCGACCGGCTCACGACTTGTGCCAGACGG 180  
QY 61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCyGlyThrGlyGlnProAla 80  
DB 181 ACCGTCCTGATGAGCGGCGCTTGCAGTCTCGATGAGGAGTGGGATACCGGACACAGCG 240  
QY 81 LeuArgValAlaAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100  
DB 241 CTGGCGGTGCGCGGACCAACCGCATCCAGATCACCGGATCACCGTACAGCCAGGTGCA 300  
QY 101 ValAlaIleAlaAlaAspCySaAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120  
DB 301 GTGGCCATCGCGCTGATTTGGCCAGCGACCGAGCTHAAGCACCGGGTGGACTTCTCG 360  
QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140  
DB 361 TGGCTGATGATGCTTCCTGCGTACCGGACATGCTTTCGACGCGCGCTGGGCGCATG 420  
QY 141 GlnSerLeuLeuGlnMetSerLeuProAspArgAlaIleArgGlnIleLeuArgValLeu 160  
DB 421 CAGTGGCTGTGGAGATGTCCAAACCGGACCGTGCATCCCGGAAATCTTCGTGATGATCTC 480  
QY 161 LysProGlyGlyIleLeuGlyValThrGluValValIleArgGluAlaGlyValGlyMet 180  
DB 481 AAACCCGGTGCATCTCTGGCGCTCACCGAGTGTCTCAACAGAGAGCGGGCGGGGATG 540

QY 181 ProValSerGlyAspArgTrpProThrGlyLeuArgIleCySLeuAlaGluGlnLeu 200  
DB 541 CCGGTGTCGGGGAGACAGTGGCCGACCGGCTTGAGATCTGCTCGGTGACCACTTCTTG 600  
QY 201 GluSerLeuArgAlaAlaGlyPheGlnIleLeuAspTrpGluAspValSerArgThr 220  
DB 601 GAATGCTGCGTGCAGCGGGGTTGAGATCTTCATTTGGAGGAGCGTGTCTCGAGGACC 660  
QY 221 ArgTyrPheMetProGlnPheAlaGluGlnLeuAlaIleHisGlnIleGlyTyrAlaAsp 240  
DB 661 CGGTACTTCAAGCCGCAATTTGCCGAGAGCTTCCTGGCACCCAGCAGGATCGCGAC 720  
QY 241 ArgTyrGlyProAlaValAlaGlyTyrAlaAlaAlaValCyAspTyrGluLysTyrAla 260  
DB 721 AGTACGGGCGGGCTGTGCGCGGTGGGCGCCCGGCGGTCTCGATTATGAGAAATATGCC 780  
QY 261 HisAspMetGlyTyrAlaIleLeuThrAlaArgLysProValGly 275  
DB 781 CACGACATGGGCTATGCGATTCTGACGCGCGGAGCGCGTCCG 825

RESULT 2

AAF88315/C  
ID AAF88315 standard; DNA; 45624 BP.

AAF88315;

28-AUG-2001 (first entry)

S. spinosa DNA fragment SEQ ID 4.

Forosamine; trimethylrhinamose; polypeptide synthase; biosynthesis;

spinosyn; polypeptide aglycone; transgenic plant; insect resistance;

macrolide; insecticidal; de.

Saccharopolyspora spinosa.

DE19957268-A1.

08-MAR-2001.

29-NOV-1999; 99DE-01057268.

27-AUG-1999; 99DE-01040596.

(FARB ) BAYER AG.

Eberz G, Moehrl V, Froede R, Velten R, Salas JA;

WPI; 2001-267102/28.

Claim 7; Page 58-74; 354pp; German.

This invention describes a novel method nucleic acid (I) and its encoded

polypeptide (II) containing at least one region that encodes an enzymatic

activity involved in biosynthesis of spinosyns. (I) are used (i) to

identify, inactive or modulate genes involved in the biosynthesis of

(ii); (ii) to generate a library of polypeptide synthases; (iii) for

adding forosamine or trimethylrhinamose to a spinosyn or polypeptide

aglycone; and (iv) for recombinant production of the corresponding

enzymes, which are used for production of (ii), their precursors or

derivatives, including production of transgenic plants that express (ii)

and thus have increased resistance to insects. (I) are also useful as

markers for sequencing of the Saccharopolyspora spinosa genome. (II) are

macroolides with insecticidal, but not antibacterial, activity, and can

also be used to raise specific antibodies, useful for identifying

expression clones in a gene bank. Cells transformed with (I) may produce

(II) at significantly increased levels or produce new derivatives of

(II). This sequence represents a genomic DNA fragment of the S. spinosa

genome which contains the coding regions for proteins involved in

forosamine and trimethylrhinamose biosynthesis

XX Sequence 45624 BP; 7933 A; 14913 C; 14940 G; 7838 T; 0 U; 0 Other:

## Alignment Scores:

Pred. No.:	Length:	Matches:
1.26e-130	45624	275
Score:	1444.00	Conservative: 0
Percent Similarity:	100.00%	Mismatches: 0
Best Local Similarity:	100.00%	Indels: 0
Query Match:	100.00%	Gaps: 0

DB: 4

US-10-069-353A-8 (1-275) x AAF88315 (1-45624)

```

QY 1 ValLeuProGlyGlyAlaPProThrSerGlnGlnValGlyGlnMetYrTrpLeuValThr 20
DB 828 GTGTTCAGGTGGCCACACCATCGACAGATGGGAGATGATGACCTGTGTACG 769
QY 21 ProLeuLeuAnsSerValAlaGlyGlyProCysAlaIleHisGlyYrTrpGluAsn 40
DB 768 CCGTGTCTGAACCTCGGTCCGGCGGCGCCCTGCGCCATCCACACCGCTACTGGAGAAC 769
QY 41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
DB 708 GACGGGGGGGCTTCCTGGACAGAGCGCCGACCGGCTCACCGACTTGTGCGCGAAACGG 649
QY 61 ThrValLeuAspGlyGlyValArgLeuLeuAspValGlyCysGlyYrTrpGlnProAla 80
DB 648 ACCGTGCTCATGGCGCGCTTCCTGCTGATGTGGGGTGGCGATGCCGGAACACACGCG 589
QY 81 LeuArgValAlaArgAspAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
DB 588 CTGCGCGCTCCGGCGACACACGGATCCATCACCGGCTACCGCTCAGCCAGGTCCAA 529
QY 101 ValAlaIleAlaIleAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
DB 528 GTGGCCATCCCGCTGATTCGGACAGCGAACCGGGAATAGACACCGGGTGAATTCTCG 469
QY 121 CysValAspAlaMetSerLeuProYrProAspAspAlaPheAspAlaAlaTPAlaMet 140
DB 468 TGGGTGATCCCATGTCTCCGCTACCCGACACATCTTCGACCGCGCTGGCCATG 409
QY 141 GlnSerLeuLeuGlnMetSerGlnProAspArgAlaIleArgGlnIleLeuArgValLeu 160
DB 408 CAGTGGTGTGGAGATGCCAACCGGACCGGTCCATCCGGAAATCTTCAGATCTC 349
QY 161 LysProGlyGlyIleLeuGlyValThrGluValValLysArgGluAlaGlyGlyMet 180
DB 348 AAACCCGGTGGCATCTCCGGCGTCAACGAGGTCTCAACGAGAGCGGGCGGGGATG 289
QY 181 ProValSerGlyAspArgTrpProThrGlyLeuArgIleCysLeuAlaGluGlnLeu 200
DB 288 CCGGTGTCCGGGGGACAGGTGGCCACCGGCTTCGAGTCTGCGCTGAGCACTTCTG 229
QY 201 GlnSerLeuArgAlaIleGlyPheGlnIleLeuAspTrpGluAspValSerSerArgThr 220
DB 228 GAATCGTGTGGAGCGGGGTTCAGATCTCATTTGGAGGAGCGTGTCTCAGGAC 169
QY 221 ArgTrpPheMetProGlnPheAlaGluGluLeuAlaIleHisGlnHisGlyIleAlaAsp 240
DB 168 CCGTACTTCATGCGCGAGTTCGCGAAGAGTCTGCTCGACACCGAGATCGGGAC 109
QY 241 ArgTrpGlyProAlaValAlaGlyYrTrpAlaAlaAlaValCysAspYrGlyLysYrTrpAla 260
DB 108 AGGATACGGGGCGGCTGTGCGCGGTGGCGCCCGCGTGTGCGATTTAGAAATATGACC 49
QY 261 HisAspMetGlyYrTrpAlaIleLeuThrAlaArgLysProValGly 275
DB 48 CACGACATGGGCTATGCGATTCTAGCGCGGAGGCGGATCGGC 4

```

RESULT 3  
AAF88312  
ID AAF88312 standard; DNA; 50000 BP.

AC AAF88312;  
XX 28-AUG-2001 (first entry)  
XX DE S. spinosa DNA fragment SEQ ID 1.  
XX Forosamine; trimethylrhamsone; polyketide synthase; biosynthesis;  
XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;  
XX macrolide; insecticidal; ds.  
XX Saccharopolyspora spinosa.  
XX DE19957268-A1.  
XX 08-MAR-2001.  
XX 29-NOV-1999; 99DE-01057268.  
XX 27-AUG-1999; 99DE-01040596.  
XX (PARB ) BAYER AG.  
XX Eberz G, Moehle V, Froede R, Velten R, Salas JA;  
XX WPI; 2001-267102/28.  
XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for  
XX recombinant production of insecticidal spinosyns and their derivatives.  
XX Claim 7; Page 14-31; 354pp; German.  
XX This invention describes a novel method nucleic acid (I) and its encoded  
XX polypeptide (II) containing at least one region that encodes an enzymatic  
XX activity involved in biosynthesis of spinosyns. (I) are used (i) to  
XX identify, inactivate or modulate genes involved in the biosynthesis of  
XX (II); (ii) to generate a library of polyketide synthases; (iii) for  
XX adding forosamine or trimethylrhamsone to a spinosyn or polyketide  
XX aglycone; and (iv) for recombinant production of the corresponding  
XX enzymes, which are used for production of (II), their precursors or  
XX derivatives, including production of transgenic plants that express (II)  
XX and thus have increased resistance to insects. (I) are also useful as  
XX markers for sequencing of the Saccharopolyspora spinosa genome. (II) are  
XX macrolides with insecticidal, but not antibacterial, activity, and can  
XX also be used to raise specific antibodies, useful for identifying  
XX expression clones in a gene bank. Cells transformed with (I) may produce  
XX (II) at significantly increased levels or produce new derivatives of  
XX (II). This sequence represents a genomic DNA fragment of the S. spinosa  
XX genome which contains the coding regions for proteins involved in  
XX forosamine, trimethylrhamsone and polyketide synthase biosynthesis

XX SQ Sequence 50000 BP; 8484 A; 16290 C; 16535 G; 8691 T; 0 U; 0 Other:

Alignment Scores:

Pred. No.:	Length:	Matches:
1.41e-130	50000	275
Score:	1444.00	Conservative: 0
Percent Similarity:	100.00%	Mismatches: 0
Best Local Similarity:	100.00%	Indels: 0
Query Match:	100.00%	Gaps: 0

DB: 4

US-10-069-353A-8 (1-275) x AAF88312 (1-50000)

```

QY 1 ValLeuProGlyGlyAlaPProThrSerGlnGlnValGlyGlnMetYrTrpLeuValThr 20
DB 44797 GTGTTCAGGTGGCCACACCATCGACAGATGGGAGATGATGACCTGTGTACG 44856
QY 21 ProLeuLeuAnsSerValAlaGlyGlyProCysAlaIleHisGlyYrTrpGluAsn 40
DB 44857 CCGTGTCTGAACCTCGGTCCGGCGGCGCCCTGCGCCATCCACACCGCTACTGGAGAAC 44916
QY 41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
DB 44917 GACGGGGGGGCTTCCTGGACAGAGCGCCGACCGGCTCACCGACTTGTGCGCGAAACGG 44976

```

QY 61 ThrValIleuAspGlyGlyValArgIleuLeuAspValGlyCysGlyThrGlyGlnProAla 80  
| | | | |  
Db 44977 ACCGTCCTGATGGCGGCGCTTGACCTGCTGATGTGGGGTGGCGTACCGGACACCGG 45036  
| | | | |  
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100  
| | | | |  
Db 45037 CTGGCGGTGGCGGCGGACCAACCGATCCAGATCACCGGATCACCGTCAGCCAGGTGCA 45096  
| | | | |  
QY 101 ValAlaIleAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120  
| | | | |  
Db 45097 GTGGCATGCGCGCTGATGCGGACGCGAAGCGGACTAAGCCACCGGTGGACTTCTCG 45156  
| | | | |  
QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrrPalMet 140  
| | | | |  
Db 45157 TCGTCGATGCCATGTCCTGCGGTCACCGGACAACTTTCAGACCGCGCTGGGCGCATG 45216  
| | | | |  
QY 141 GlnSerLeuLeuGluMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160  
| | | | |  
Db 45217 CAGTCGCTGTGGAGATGTCGAACCGGACCGTCGATCCGGGAAATCCTTCGAGTACTC 45276  
| | | | |  
QY 161 LysProGlyGlyIleLeuGlyValThrGlyValIleValArgGluAlaGlyGlyMet 180  
| | | | |  
Db 45277 AACCCGGTGCACTCTCGGCGTCACCGAGTGTCAACGAGAGCGGCGCGGCGATG 45336  
| | | | |  
QY 181 ProValSerGlyAspArgTrrProThrGlyLeuArgIleCysLeuAlaGluGlnLeu 200  
| | | | |  
Db 45337 CCGGTGCGGGGACAGGTGGCGACCGGCTTCGGATTTGCTGCTAGCACTTCTG 45396  
| | | | |  
QY 201 GluSerLeuArgAlaAlaGlyPheGluIleLeuAspTrrGluAspValSerSerArgThr 220  
| | | | |  
Db 45397 GAATGCGCTGCGGACGCGGGGTCGAGATCTCGATTGGGAGACGCTGCTCGAGGACC 45456  
| | | | |  
QY 221 ArgTyrPheMetProGlnPheAlaGluGluLeuAlaHisGlyIleAlaAsp 240  
| | | | |  
Db 45457 CCGTACTTATCGCGGATTCGCGGAGAGCTGCTGCGCACGACGAGCGGATCGGGAC 45516  
| | | | |  
QY 241 ArgTyrGlyProAlaValAlaGlyTrrPalAlaAlaValCysAspTyrGluYsTrrAla 260  
| | | | |  
Db 45517 AGGTACGGGCGGCGTGTGCGGCGGCGCGCGGCTGTGCGATTATGAGAAATATGCC 45576  
| | | | |  
QY 261 HisAspMetGlyTrrAlaIleLeuThrAlaArgLysProValGly 275  
| | | | |  
Db 45577 CACGACATGGCTATGCGATCTGACGCGCGGAAACCGCTCGC 45621  
| | | | |  
RESULT 4  
AAZ21501  
ID AAZ21501 standard; DNA; 80161 BP.  
XX  
XX AAZ21501;  
AC  
XX  
XX 01-DEC-1999 (first entry)  
DT  
XX  
XX  
DE DNA fragment of Saccharopolyspora spinosa containing biosynthetic genes.  
XX  
XX Spinosyn biosynthetic enzyme; open reading frame; ORF;  
KW insecticidal microides; arachnid; nematode; insect; polyketide;  
KW polyketide synthase; PKS; extender module; initiator module;  
KW acyl transferase domain; AT; acyl carrier protein; ACP;  
KW beta-ketoadenylase domain; KS; KR; dehydratase domain; DH;  
KW enoyl reductase domain; ER; beta-ketoreductase; insecticide; ss.  
OS  
XX Saccharopolyspora spinosa.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS complement(1135..1971)  
FT /tag= a  
FT /product= "ORF16"  
FT /note= "Protein involved in transcription control"  
FT 2024..2791  
FT /tag= b  
FT /product= "ORF15"  
FT /note= "Keto acyl reductase"

FT CDS complement(3416..4165)  
FT /tag= c  
FT /product= "spnS"  
FT /note= "Spinosyn biosynthesis protein S"  
FT complement(4168..5325)  
FT /tag= d  
FT /product= "spnR"  
FT /note= "Spinosyn biosynthesis protein R"  
FT complement(5363..6751)  
FT /tag= e  
FT /product= "spnQ"  
FT /note= "Spinosyn biosynthesis protein Q"  
FT 7083..8450  
FT /tag= f  
FT /product= "spnP"  
FT /note= "Spinosyn biosynthesis protein P"  
FT 8967..10427  
FT /tag= g  
FT /product= "spnO"  
FT /note= "Spinosyn biosynthesis protein O"  
FT 10436..11434  
FT /tag= h  
FT /product= "spnN"  
FT /note= "Spinosyn biosynthesis protein N"  
FT complement(11530..12492)  
FT /tag= i  
FT /product= "spnM"  
FT /note= "Spinosyn biosynthesis protein M"  
FT complement(12696..13547)  
FT /tag= j  
FT /product= "spnL"  
FT /note= "Spinosyn biosynthesis protein L"  
FT 13592..14785  
FT /tag= k  
FT /product= "spnK"  
FT /note= "Spinosyn biosynthesis protein K"  
FT complement(14799..16418)  
FT /tag= l  
FT /product= "spnJ"  
FT /note= "Spinosyn biosynthesis protein J"  
FT 16556..17743  
FT /tag= m  
FT /product= "spnI"  
FT /note= "Spinosyn biosynthesis protein I"  
FT complement(17749..18501)  
FT /tag= n  
FT /product= "spnH"  
FT /note= "Spinosyn biosynthesis protein H"  
FT complement(18541..19713)  
FT /tag= o  
FT /product= "spnG"  
FT /note= "Spinosyn biosynthesis protein G"  
FT 20168..20995  
FT /tag= p  
FT /product= "spnF"  
FT /note= "Spinosyn biosynthesis protein F"  
FT 21111..28898  
FT /tag= q  
FT /product= "spnA"  
FT /note= "Spinosyn biosynthesis protein A"  
FT /transl\_except= (Pos:26940..26942, aa:Pro)  
FT 21126..22379  
FT /tag= r  
FT /note= "Beta-ketoadenylase domain (KS): part of the  
FT initiator module"  
FT 22692..23669  
FT /tag= s  
FT /note= "Acyl transferase domain (AT1): part of the  
FT initiator module"  
FT 23793..24041  
FT /tag= t  
FT /note= "Acyl carrier protein domain (ACPI): part of the  
FT initiator module"

```
FT misc_feature 24102..25349
FT /tag= u
FT /note= "Beta-ketosynthase domain (KS1) : part of extender
FT module 1"
FT 25683..26684
FT /tag= v
FT /note= "Acyl transferase domain (AT1) : part of extender
FT module 1"
FT 27582..28121
FT /tag= w
FT /note= "Beta-ketoreductase domain (KR1) : part of extender
FT module 1"
FT 28404..28649
FT /tag= x
FT /note= "Acyl carrier protein domain (ACP1) : part of
FT extender module 1"
FT 29024..30295
FT /tag= z
FT /note= "Beta-ketosynthase domain (KS2) : part of extender
FT module 2"
FT 30629..31621
FT /tag= aa
FT /note= "Acyl transferase domain (AT2) : part of extender
FT module 2"
FT 31697..32254
FT /tag= ab
FT /note= "Dehydratase domain (DH2) : part of extender module
FT 2"
FT 33035..34072
FT /tag= ac
FT /note= "Enoyl reductase domain (ER2) : part of extender
FT module 2"
FT 34082..34621
FT /tag= ad
FT /note= "Beta-ketoreductase domain (KR2) : part of extender
FT module 2"
FT 34886..350295
FT /tag= ae
FT /note= "Acyl carrier protein domain (ACP2) : part of
FT extender module 2"
FT 35419..44931
FT /tag= af
FT /product= "spnc"
FT /note= "Spinosyn biosynthesis protein C"
FT 35518..36786
FT /tag= ag
FT /note= "Beta-ketosynthase domain (KS3) : part of extender
FT module 3"
FT 37108..38097
FT /tag= ah
FT /note= "Acyl transferase domain (AT3) : part of extender
FT module 3"
FT 38916..35374
FT /tag= y
FT /product= "spnc"
FT /note= "Spinosyn biosynthesis protein C"
FT 38992..39528
FT /tag= ai
FT /note= "Beta-ketoreductase domain (KR3) : part of extender
FT module 3"
FT 39790..40035
FT /tag= aj
FT /note= "Acyl carrier protein domain (ACP3) : part of
FT extender module 3"
FT 40102..41373
FT /tag= ak
FT /note= "Beta-ketosynthase domain (KS4) : part of extender
FT module 4"
FT 41713..42705
FT /tag= al
FT /note= "Acyl transferase domain (AT4) : part of extender
FT module 4"
FT 43615..44157
FT misc_feature
```

```
FT /tag= am
FT /note= "Beta-ketoreductase domain (KR4) : part of extender
FT module 4"
FT 44431..44676
FT /tag= an
FT /note= "Acyl carrier protein domain (ACP4) : part of
FT extender module 4"
FT 44966..59752
FT /tag= ao
FT /product= "spnd"
FT /note= "Spinosyn biosynthesis protein D"
FT 45077..46348
FT /tag= ap
FT /note= "Beta-ketosynthase domain (KS5) : part of extender
FT module 5"
FT 46691..47674
FT /tag= aq
FT /note= "Acyl transferase domain (AT5) : part of extender
FT module 5"
FT 47753..48310
FT /tag= ar
FT /note= "Dehydratase domain (DH5) : part of extender module
FT 5"
FT 49226..49771
FT /tag= as
FT /note= "Beta-ketoreductase domain (KR5) : part of extender
FT module 5"
FT 50009..50254
FT misc_feature

Alignment Scores:
Pred. No.: 2,55e-130 Length: 80161
Score: 1444.00 Matches: 275
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-10-069-353a-8 (1-275) x AA21501 (1-80161)

QY 1 ValLeuProGlyValAlaProThSerGInGInValGInMeTyRAspLeuValThr 20
DB GTGTGTCAGTGGCCACCAACATCGACAGTGGGAGATGATGACCTGGTCACG 20227
QY 21 ProLeuLeuAsnSerValAlaGlyIProCVaAlaIleHisIaGlyTYrTrpGluAsn 40
DB CCGTGGTGAACCTGGTCGGGGGCGCCCTGGCCATCCACACGACTGAGAGAAC 20287
QY 41 AspGlyARgAlaSerTrpGInGInAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
DB GACGGGGGGGCTTCCTGGACGACGCGCGACCGGCTACCGACTGTGCGCGAACG 20347
QY 61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCYsGlyThyGlyInProAla 80
DB ACCGTGCTGATGGCGGCGTTCGACTGCTGATGGGGTGGCGGTACCGGACACGCG 20407
QY 81 LeuArgValAlaArgAspAsnAlaIleGInIleThrGlyYrLeThrValSerGInValGln 100
DB CTGGCGGTCCGGCGCAACGCGATCCAGATCCAGGCACTACCGGACAGGTGCA 20467
QY 101 ValAlaIleAlaAlaAspCYsAlaArgGluArgGlyLeuSerHisArgValaAspPheSer 120
DB GTGGCCATCGCCGCTGATTCGACGACGGAACGCGACTAAGCAGCGGGTGACTTCG 20527
QY 121 CysValaAspAlaMetSerLeuProTYrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
DB TGGCTGATCCATGTGCTCCGCGGTACCGGCAAAATGCTTTCAGCGCGCTGGCCATG 20587
QY 141 GluSerLeuLeuGluMetSerGluProAspArgAlaIleArgGluIleLeuArgValaLeu 160
DB CAGTCGTGTTGAGATGTCGAACCGGACGATCCGAGAAATCTTTCAGATGACTC 20647
QY 161 LysProGlyIleLeuGlyValThrGluValaValaLysArgGluAlaGlyIleGlyMet 180
DB |||||
```

Db 20648 AAACCGGATGCGATCTCGGCGTCAACGAGTCTCAAAAGAAAGCGGCGCGGATG 20707  
Qy 181 ProValSerGlyAspArgTrpProThrGlyLeuArgIleCysLeuAlaGluGlnLeu 200  
Db 20708 CCGGTCCTCCGGGACAGGTGGCCGACCGGCTTCGGATTCGCTGCTGAGCACTTCTG 20767  
Qy 201 GluSerLeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220  
Db 20768 GAATCGCTCGTGCAGCGGGGTTCCAGATCTCGATTGGAGAGACGTGTCTGAGAGACC 20827  
Qy 221 ArgTrpPheMetProGlnPheAlaGluGlnLeuAlaAlaHisGlyIleAlaAsp 240  
Db 20828 CCGTACTTATGATCCGAGTTCGCCGAAGAGCTCGCTGCCACAGCACGGGATCCGGAGC 20887  
Qy 241 ArgTrpGlyProAlaValAlaGlyTrpAlaAlaAlaValCysAspTrpGluIleArgIle 260  
Db 20888 AGGTACGGGCGCGCTCTCCCGCTGGCGCCGCCGCTGTGCGATTATGAGAAATATGCC 20947  
Qy 261 HisAspMetGlyTrpAlaIleLeuThrAlaArgIleAspProValGly 275  
Db 20948 CACGACATGGCTATGCGATCTCGACGCGCGGAAGCCGCTCGGC 20992  
RESULT 5  
ABV75558/c  
ID ABV75558 standard; DNA; 36538 BP.  
XX  
XX ABV75558;  
XX  
DT 22-JAN-2003 (first entry)  
XX  
XX Saccharopolyspora butenyl-spinosyn biosynthetic gene cluster 2.  
XX  
XX Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrolide;  
KW metabolite; spinosyn; gene; ds.  
XX  
XX Saccharopolyspora sp.  
OS  
XX  
XX Key Location/Qualifiers  
FH CDS Complement(114..938)  
FT /tag= a  
FT /product= "busF"  
FT /note= "No start codon given"  
FT 1389..2561  
FT /tag= b  
FT /product= "busG"  
FT 2601..3353  
FT /tag= c  
FT /product= "busH"  
FT complement(3359..4546)  
FT /tag= d  
FT /product= "busI"  
FT 4684..6303  
FT /tag= e  
FT /product= "busJ"  
FT 6317..7510  
FT /tag= f  
FT /product= "busK"  
FT 7555..8406  
FT /tag= h  
FT /product= "busM"  
FT /note= "No start codon given"  
FT 8640..9572  
FT /tag= h  
FT /product= "busM"  
FT /note= "No start codon given"  
FT complement(9668..10666)  
FT /tag= i  
FT /product= "busN"  
FT complement(10675..12135)  
FT /tag= j  
FT /product= "busO"  
FT complement(12864..14177)  
FT /tag= k  
FT CDS

FT /product= "busP"  
FT 14627..15970  
FT /tag= l  
FT /product= "busQ"  
FT /note= "No start codon given"  
FT 16008..17144  
FT /tag= m  
FT /product= "busR"  
FT /note= "No start codon given"  
FT 17168..17917  
FT /tag= n  
FT /product= "busS"  
FT complement(18520..19932)  
FT /tag= o  
FT /product= "ORF LI"  
FT complement(19978..20488)  
FT /tag= p  
FT /product= "ORF LII"  
FT /note= "No start codon given"  
FT complement(20536..21033)  
FT /tag= q  
FT /product= "ORF LIII"  
FT /note= "No start codon given"  
FT 21179..21925  
FT /tag= r  
FT /product= "ORF LIV"  
FT complement(22671..23453)  
FT /tag= s  
FT /product= "ORF LVI"  
FT complement(23687..24886)  
FT /tag= t  
FT /product= "ORF LVII"  
FT complement(26177..26923)  
FT /tag= u  
FT /product= "ORF LVIII"  
FT /note= "No start codon given"  
FT 27646..28476  
FT /tag= v  
FT /product= "ORF LIX"  
FT /note= "No start codon given"  
FT  
FN WO200279477-A2.  
XX  
XX PD 10-OCT-2002.  
XX  
XX 28-MAR-2002; 2002WO-US009968.  
XX  
XX PR 30-MAR-2001; 2001US-0280175P.  
XX  
XX (DOWC ) DOW AGROSCIENCES LLC.  
XX  
XX PA Hahn DR, Jackson JD, Bullard BS, Guevafson GD, Waldron C;  
PI Mitchell JC;  
XX  
XX WIPI; 2003-058434/05.  
DR P-PsDB; ABP57683, ABP57684, ABP57685, ABP57686, ABP57687, ABP57688,  
DR ABP57689, ABP57690, ABP57691, ABP57692, ABP57693, ABP57694, ABP57695,  
DR ABP57696, ABP57697, ABP57698, ABP57699, ABP57700, ABP57701, ABP57702,  
DR ABP57703, ABP57704.  
XX  
XX New butenyl-spinosyn biosynthetic genes, useful for increasing the  
FT production of butenyl-spinosyn insecticidal macrolides, or for changing  
FT the metabolites or products produced by spinosyn-producing  
FT microorganisms.  
XX  
XX Claim 2; Page 99-119; 218pp; English.  
XX  
XX The invention relates to a novel DNA molecule comprising a DNA sequence  
CC that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn  
CC PKS (polyketide synthetase) domain, or a spinosyn PKS module. The butenyl-  
CC spinosyn biosynthetic genes are useful for increasing the production of  
CC butenyl-spinosyn insecticidal macrolides. The genes are also useful for  
CC changing the metabolites or products produced by spinosyn-producing

CC microorganisms. The present sequence represents a DNA molecule encoding  
CC butenyl-epinosyn biosynthetic enzymes

XX Sequence 36538 BP; 6867 A; 12266 C; 11182 G; 6223 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4 93e-118	Length:	36538
Score:	1315.00	Matches:	252
Percent Similarity:	94.18%	Conservative:	7
Best Local Similarity:	91.64%	Mismatches:	16
Query Match:	91.07%	Indels:	0
DB:	10	Gaps:	0

US-10-069-353a-8 (1-275) x ABV75558 (1-36538)

```

QY 1 ValLeuProGlyGlyValAlaProThrsSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20
   |||||
DB 938 GTGTTGCCAGGTGGCGCTGCCACATCGCAGCAGGTGGGCAATGATGACCTGGTCACC 879

QY 21 ProLeuLeuAsnSerValAlaGlyGlyProGlyAlaIleHisGlyTyrTrpGluAsn 40
   |||||
DB 878 CCGTGTCTGAATTCGGTCCGGCGCGCCCTGCGCATCCACACGCTACTGGAGAAC 819

QY 41 AArgGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
   |||||
DB 818 GACGGGGGACCTTCCTGGCAGCAGCGCGCCGACCGGCTCAACGATCTTGTCCCGCAACG 759

QY 61 ThrValLeuAspGlyGlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80
   |||||
DB 758 ACCGGCGCTGATGGCGGCAATGACCTGCGGATGTGGGGTGGCGGACCGGACCAACGACG 699

QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
   |||||
DB 698 CTGCGCGTGGCGCGCAACGCGATCCGCGATCAGCGGCACTCACGGTCAAGCAGGTGCA 639

QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
   |||||
DB 638 GCTGCGCATCGCTTGAATTTGGCGCCCGGAAACGCGACTAAGCCACGAGTGACTCTCG 579

QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
   |||||
DB 578 TGGTGCAGCCCATGTCCTCCGCGCATCCGCAATCGTTGACCGCTGCTGGGCGATC 519

QY 141 GlnSerLeuLeuGlnLeuSerGlnProAspArgAlaIleArgGluIleLeuArgValLeu 160
   |||||
DB 518 CAGTCCGCTGTGAGATGTCGGAACCGGACCGTCCCATCCGGGAAATCGTTGAGTGTCTC 459

QY 161 LysProGlyGlyIleLeuGlyValThrGluValValLysArgGluAlaGlyGlyMet 180
   |||||
DB 458 AAACCCGGTGGCATCTCCGAGTCAACGAGGTGCTCAACGAGAAATAGCAGCGGATC 399

QY 181 ProValSerGlyAspArgTrpProThrGlyLeuArgIleCysLeuAlaGlnLeu 200
   |||||
DB 398 CCGGTGTCCGGGACATGTGGCCGACCGGCTCCGATCTGCTGGTGAACAATTCTG 339

QY 201 GluSerLeuArgAlaAlaGlyPheGlnIleLeuAspTrpGluAspValSerSerArgThr 220
   |||||
DB 338 GAATCGCTGTGTCAGCGGGTTCGAAATCTCCCTGTGTGACGACGTCGTGTCGCGGACC 279

QY 221 ArgTyrPheMetProGlnPheAlaGluGlnLeuAlaAlaHisGlnHisGlyIleAlaAsp 240
   |||||
DB 278 CGGACTTCAATGCCGCGAGTCCGGAACCGCTCCCTCGGACCGACCGGATCGCGGAA 219

QY 241 ArgTyrGlyProAlaValAlaGlyTrpAlaAlaAlaValCysAspTyrGluLysTyrAla 260
   |||||
DB 218 AGGTACGGGGCGGCTGTGCGCGACTGGGGCGCGCGGTCTGCGATTAAGAAATACGCC 159

QY 261 HisAspMetGlyTyrAlaIleLeuThrAlaArgLysProValGly 275
   |||||
DB 158 GACGACATGGGCTATGCGATTTTGACGCGCGGAAACCGGTCGCGC 114
```

RESULT 6  
ABV75558

```

ID ABV75558 standard; DNA; 36538 BP.
XX
AC ABV75558;
XX
DT 22-JAN-2003 (first entry)
XX
DE Saccharopolyspora butenyl-epinosyn biosynthetic gene cluster 2.
XX
KM Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macroide;
XX metabolite; spinosyn gene; ds.
XX
OS Saccharopolyspora sp.
XX
FH Key
FT CDS
   Location/Qualifiers
   /tag= a
   /product= "bueF"
   /note= "No start codon given"
   1389..2561
   /tag= b
   /product= "bueG"
   2601..3353
   /tag= c
   /product= "bueH"
   /note= "complement(3359..4546)"
   /tag= d
   /product= "bueI"
   4684..6303
   /tag= e
   /product= "bueJ"
   6317..7510
   /tag= f
   /product= "bueK"
   7555..8406
   /tag= g
   /product= "bueL"
   /note= "No start codon given"
   8640..9572
   /tag= h
   /product= "bueM"
   /note= "No start codon given"
   complement(10675..12135)
   /tag= j
   /product= "bueO"
   /note= "complement(12864..14177)"
   /tag= k
   /product= "bueP"
   14627..15970
   /tag= l
   /product= "bueQ"
   /note= "No start codon given"
   15008..17144
   /tag= m
   /product= "bueR"
   /note= "No start codon given"
   17168..17917
   /tag= n
   /product= "bueS"
   /note= "complement(18520..19932)"
   /tag= o
   /product= "ORF LI"
   /complement(19978..20488)
   /tag= p
   /product= "ORF LI1"
   /note= "No start codon given"
   complement(20536..21033)
   /tag= q
   /product= "ORF LI11"
   /note= "No start codon given"
   21179..21925
   CDS
```

```
FT FT      /*tag= r
FT FT      /product= "ORF LIV"
FT FT      complement(22671..23453)
FT FT      CDS
FT FT      /*tag= s
FT FT      /product= "ORF LVI"
FT FT      complement(23687..24886)
FT FT      CDS
FT FT      /*tag= t
FT FT      /product= "ORF LVII"
FT FT      complement(26177..26923)
FT FT      CDS
FT FT      /*tag= u
FT FT      /product= "ORF LVIII"
FT FT      /note= "No start codon given"
FT FT      27646..28476
FT FT      /*tag= v
FT FT      /product= "ORF LIX"
FT FT      /note= "No start codon given"
XX XX      WO200279477-A2.
XX XX      PD
XX XX      10-OCT-2002.
XX XX      PF
XX XX      28-MAR-2002; 2002WO-US009968.
XX XX      PR
XX XX      30-MAR-2001; 2001US-0280175P.
XX XX      PA
XX XX      (DOWC ) DOW AGROSCIENCES LLC.
XX XX      PI
XX XX      Hahn DR, Jackson JD, Bullard BS, Gustafson GD, Waldron C;
XX XX      Mitchell JC;
XX XX      WPI; 2003-058434/05.
XX XX      DR
XX XX      P-PEDB; ABP57683, ABP57684, ABP57685, ABP57686, ABP57687, ABP57688,
XX XX      DR
XX XX      ABP57689, ABP57690, ABP57691, ABP57692, ABP57693, ABP57694, ABP57695,
XX XX      DR
XX XX      ABP57696, ABP57697, ABP57698, ABP57699, ABP57700, ABP57701, ABP57702,
XX XX      DR
XX XX      ABP57703, ABP57704.
XX XX      PT
XX XX      New butenyl-spinosyn biosynthetic genes, useful for increasing the
XX XX      PT
XX XX      production of butenyl-spinosyn insecticidal macrolides, or for changing
XX XX      PT
XX XX      the metabolites or products produced by spinosyn-producing
XX XX      PT
XX XX      microorganisms.
XX XX      PS
XX XX      Claim 2; Page 99-119; 218pp; English.
XX XX      CC
XX XX      The invention relates to a novel DNA molecule comprising a DNA sequence
XX XX      CC
XX XX      that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn
XX XX      CC
XX XX      PKS (polyketide synthetase) domain, or a spinosyn PKS module. The butenyl-
XX XX      CC
XX XX      spinosyn biosynthetic genes are useful for increasing the production of
XX XX      CC
XX XX      butenyl-spinosyn insecticidal macrolides. The genes are also useful for
XX XX      CC
XX XX      changing the metabolites or products produced by spinosyn-producing
XX XX      CC
XX XX      microorganisms. The present sequence represents a DNA molecule encoding
XX XX      CC
XX XX      butenyl-spinosyn biosynthetic enzymes
XX XX      SQ
XX XX      Sequence 36538 BP; 6867 A; 12266 C; 11182 G; 6223 T; 0 U; 0 Other;
XX XX
XX XX      Alignment Scores:
XX XX      Pred. No.: 9,42e-32 Length: 36538
XX XX      Score: 439.50 Matches: 105
XX XX      Percent Similarity: 52.36% Conservat: 50
XX XX      Best Local Similarity: 35.47% Mismatches: 100
XX XX      Query Match: 30.44% Indels: 41
XX XX      DB: 10 Gaps: 8
XX XX
XX XX      US-10-069-353A-8 (1-275) x ABV75558 (1-36538)
XX XX
XX XX      QY 7 ProthrserGlnGlnValGlyGlnMetYrAspLeuValThrProLeuLeuAsnSerVal 26
XX XX      DB 7540 CCAAGCGGAGTACAGTGCATCTTCGAC-----GCCGTG 7578
XX XX
XX XX      QY 27 AlaGlyGlyProCysAlaIleHisIsglyTyrTrp-----GluAsnAsp 41
XX XX      DB 7579 GCGCAGAGGCGC---GCCCTGCACCAACGATTACTGGCGCGCGGCGGATTCGGAGGATGCC 7635
XX XX
XX XX      QY 42 GlyArgAlaSerTyrGlnGlnAlaIleAspArgLeuThrAspLeuValAlaGluArgThr 61
```

```
DB 7636 GGGGCGACACCTTGGTGGAGCGCTGCCGACCACTGCAGACTCTTTCATGCAGACAGGCC 7695
QY 62 ValLeuAspGlyGlyValArgLeuAspValGlyCysGlyThrGlyGlnProAlaLeu 81
DB 7686 GCGCTCCCGCCCGGAGCGCACCTGTTCCAGCTGGCGCTTGCCAAAGGGCAGCCCGTACGTC 7755
QY 82 ArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGlnVal 101
DB 7756 CGCGCGGACGACCAAGAGCGTTGAGTACCGGAAATCACCGTAAACGCCGACATCTTC 7815
QY 102 AlaIleAlaIleAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSerCys 121
DB 7816 GCCCGCGTACCAAGCTGCCCAACGACGCGACTGGCGCAAGTCTTCGGTTGCATCTGA 7875
QY 122 ValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaIleThrAlaMetGln 141
DB 7876 GTCCAGCGCGCGCGCTCCCTACCCGGAAGGTTCTTTCACGCGCGCATGGGCGATGCGAG 7935
QY 142 SerLeuLeuGluMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeuLys 161
DB 7936 TCCGTGATCAGATGTCGACCAAGCTGCCGATCCGCGAGTCCACCGAATCTCGAA 7995
QY 162 ProGlyGlyIleLeuGlyValThrGluValValIysArgGluAlaGlyGlyMetPro 181
DB 7996 CCCGCGCGCAGTTCGTCCTCGGGAGCATCACTCGT-----GCTCGACTCCG 8046
QY 182 ValSerGlyAspArgTyrProThrGlyLeuArgIleCysIleAlaGluGlnLeuGlu 201
DB 8047 GAAAGTACCGCGGGGTGGACCGCGCACGACCGCCATCCTGAACAGCTTACCGCG 8106
QY 202 SerLeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThrArg 221
DB 8107 CTGTAACGGAAGCGCGGTTTGAAGTTTTCGAAGTCAACCGACCTCAGCGCGCACACGA 8166
QY 222 TyrPheMetProGlnIlePheAlaGluGlnLeu----- 231
DB 8167 TGCATGTCCTCTGTATGTCGACGAGTGTCTCCGGAACTCGATGAGCTCGCGCGCTC 8226
QY 232 -----AlaAlaHisGlnHis-----GlyIleAlaAspArgTyr 242
DB 8227 GAGCCTGCGGTGTCGCGACCTCCAGCAACGACTTGGAGACATCGCGGAGAGCAC 8286
QY 243 GlyProAlaValAlaGlyTyrPalaAlaAlaValCysAspTyrGluLys-----Tyr 259
DB 8287 GACCGCGGACCAAGCGACGCTGATCGCCCGGTCCGGAATACCGGAACATCCCGATTAC 8346
QY 260 AlaHisAsp-----MetGlyTyrAlaIleLeuThrAlaArgLys 272
DB 8347 GCCAGAAACGAGAAAGCATGGTTTCATGCTCTGCGAGGCGGAAAG 8394
XX XX
XX XX      RESULT 7
XX XX      AA080524/C
XX XX      ID AA080524 standard; DNA; 1410 BP.
XX XX      AC
XX XX      AA080524;
XX XX      DT 10-APR-1995 (first entry)
XX XX      DE Oxidoreducing avermectin DNA from S. avermectilis ATCC31271.
XX XX      KW Modified; oxidoreducing; avermectin; S. avermectilis; wild type; disabled;
XX XX      OS Streptomyces avermectilis.
XX XX      FH Key Location/Qualifiers
XX XX      FT misc_feature 627..1075
XX XX      FT /*tag= a
XX XX      FT /note= "Avermectin coding sequence"
XX XX      PN JP06189774-A.
```





## Alignment Scores:

Pred. No.:	1,069-33	Length:	846
Score:	438.50	Matches:	106
Percent Similarity:	56.10%	Conservative:	55
Best Local Similarity:	36.93%	Mismatches:	95
Query Match:	30.37%	Indels:	31
DB:	10	Gaps:	9

US-10-069-353a-8 (1-275) x AAD54227 (1-846)

```

QY      2 LeuProGlyGlyAlaProThrSerGlnGlnValGlyGlnMetTyraSpleuValThrPro 21
        ::::::::::::::::::::
DB      34 GTCCCGGATCTCCGCGCCCGCCGGAAGAGTGGAGCACTCTACGACCGGCTCACCGGCA 93
QY      22 LeuLeuAsnSerValAlaGlyGlyProCysAlaIleHisGlyTyrThrGlu---Asn 40
        |||
DB      94 CTGAGACACCGAAGCGCGCGCGGCGC---AGCTTCACCTCGGCTACTGGAGACTGCGAC 147
QY      41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
        ::::::::::::::::::::
DB      148 GACACAGCAACCGCCGCTCGTAGAAGGCGCGGACCGGCTACCGACACGATGACCGACCGC 207
QY      61 ThrValLeuAspGlyGlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80
        ::::::::::::::::::::
DB      208 CTGCGGATCGACAGGAGGACAGCGGGCTCTCGACGTGCGCTGGGAGTCCGCGACCGCGCC 267
QY      81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
        ::::::::::::::::::::
DB      268 ATGCGGATCGCGCGCGCGGACCGCGCCCATGTGTCAGCGGATCGCATGCAAGACGACG 327
QY      101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
        ::::::::::::::::::::
DB      328 ATCGCGCGCGGACACCGGCTCGCGGAGGCGCGCGCTGAGCGACCGGCTGAGATTCCG 387
QY      121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaThrAlaMet 140
        |||
DB      388 CACGCGGACGCGCATGAACTGCCCTTCCCGACGACTCTTCGACGCGCGCATCGCCATC 447
QY      141 GlnSerLeuGlnMetSerGluProAspArgAlaIleArgGlnIleLeuArgValLeu 160
        ::::::::::::::::::::
DB      448 GAGTGATCTTCCACATGCGCGGACCGGAGCGGGTCTCGCGGAGATCGCGCGCTACTCG 507
QY      161 LysProGlyGlyIleLeuGlyValThrGluValIleLysArgGluAlaGlyGlyMet 180
        ::::::::::::::::::::
DB      508 CGCGCGGCGCGCGGCTGCTCTCAACCATCTTCGAGCGCGGCGC----- 552
QY      181 ProValSerGlyAspArgTrpProThrGly-----LeuArg 192
        ::::::::::::::::::::
DB      553 CCGGTCGCCCGGACGAGACGCGCGGTGACCGGCTCTCGGACTTCATCATGTGACG 612
QY      193 IleCysLeuAlaGlnIleLeuGlnIleSerLeuArgAlaAlaGlyPhe-----Glu 209
        ::::::::::::::::::::
DB      613 CTGCGCGCGCGCGGACGATACGCTGCGCATGCTGCGCGAGCGCGCTCGCTCGAG 672
QY      210 IleLeuAspTrpGluAspValSerSerArgThrArgTyr---PheMetProGlnPheAla 228
        ::::::::::::::::::::
DB      673 CTCTCGACATACCGAGAGCGGTGCTGACGACTTCGAGCAATAGACGAGGCGCTCC 732
QY      229 GluGluLeu---AlaAlaHisGlnHisGlyIleAlaAspArgTyrGlyProAlaValAla 247
        ::::::::::::::::::::
DB      733 CAGGAGATGACGACCGCTTTCGACGACGAGCGAGCAAGTTCAGGCC----- 783
QY      248 GlyTrpAlaAlaAlaValCysAspTyrGluLysTyrAlaHisAspMetGlyTyrAlaIle 267
        ::::::::::::::::::::
DB      784 -----GCTTCATGATGACCTCGACGCAATTC-----GCTCGGTTCTG 822
QY      268 LeuThrAlaArgLysProVal 274
        |||
DB      823 CTGACCGCGCAAAACCCCTC 843

```

RESULT 9  
AAD54217

```

ID      AAD54217 standard; DNA; 52101 BP.
XX
AC      AAD54217;
XX
DT      17-JUN-2003 (first entry)
XX
DE      Streptomyces platensis subspecies rosaceus dorrigocin DNA.
XX
KM      Polyketide biosynthesis; dorrigocin; DORR; laccimidomycin; LACT; gene;
XX      ds.
XX
XX      Streptomyces platensis.
XX
OS      Location/Qualifiers
FH      complement(67..370)
FH      CDS
FT      /tag= a
FT      /product= "ORF1 protein"
FT      /tag= b
FT      /product= "ORF2 protein"
FT      /note= "No start codon"
FT      /partial
FT      CDS
FT      /tag= c
FT      /product= "ORF3 protein"
FT      /tag= d
FT      /product= "ORF4 protein"
FT      /tag= e
FT      /product= "ORF5 protein"
FT      /tag= f
FT      /product= "ORF6 protein"
FT      /note= "No start codon"
FT      /partial
FT      /tag= g
FT      /product= "ORF7 protein"
FT      /note= "No start codon"
FT      /partial
FT      /tag= h
FT      /product= "ORF8 protein"
FT      /tag= i
FT      /product= "ORF9 protein"
FT      /tag= j
FT      /product= "ORF10 protein"
XX
XX      MO200288176-A2.
XX
XX      PD
XX      07-NOV-2002.
XX
XX      PF
XX      26-APR-2002; 2002MO-CA000591.
XX
XX      PR
XX      26-APR-2001; 2001US-0286346P.
XX
XX      PA
XX      (ECOP-) ECOP1A BIOSCIENCES INC.
XX
XX      PI
XX      Farnet CM, Zazopoulos E, Staffa A, Yang X;
XX
XX      DR
XX      WPI, 2003-201222/19.
XX      DR
XX      P-PSDB; AAE35484, AAE35485, AAE35486, AAE35487, AAE35488, AAE35489,
XX      AAE35490, AAE35491, AAE35492, AAE35493.
XX
XX      Novel isolated or purified polypeptide involved in biosynthesis of
XX      polyketide dorrigocin or polyketide laccimidomycin, useful for preparing
XX      dorrigocin or laccimidomycin.
XX
XX      Claim 1; Page 85-113; 312pp; English.
XX

```



```

Best Local Similarity: 36.00%      Mismatches: 96
Query Match: 30.16%      Indels: 49
DB: 4      Gaps: 11
US-10-069-353a-8 (1-275) x AAF88315 (1-900)

QY 7 ProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThrProLeuLeuAsnSerVal 26
DB 34 CCGACCGCGGATCAGGTGAGTCCATCTTCGATGCGTTG-----GGC 75
QY 27 AlaGlyGlyProCysAlaIleHisIleGlyTyrTrp-----GluAsnAsp 41
DB 76 CACGGGCGTCCC-----CTGCACCACGGTACTGGCGGGCGGGGTATCCGAGAGATGCC 129
QY 42 GlyArgAlaSerTrpGlnGlnIleAlaAspArgLeuThrAspLeuValAlaGluArgThr 61
DB 130 GGTGGCACACCGCTGGTGGATGCTGCCACCACTGACCGACCTGTTCATCACAAGGCC 189
QY 62 ValLeuAspGlyGlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAlaLeu 81
DB 190 GCGCTCCGTCGGGAGCGCAGCTGTTCGACTGGGCTGGCGCAATGGGAGCCCGTAGTC 249
QY 82 ArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGlnVal 101
DB 250 CGTGGCGCATGGCCAGCGCGCGCTTGAGTACCGGAAATCACCGTGAACGCCAGCATCTTC 309
QY 102 AlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSerCys 121
DB 310 GCCCGCCGACACAGGCTGCCCAAGACCGGACCTGGCGGAGCTTGAGTTGCATCTA 369
QY 122 ValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMetGln 141
DB 370 GTCCGCGCGCCACACTGCTCCCTACCCGCGCGGCTTTCTTCAGCGCGCATGGCGATGAG 429
QY 142 SerLeuLeuGluMetSerGluProAspArgAlaIleArgGlyIleLeuArgValLeuArg 161
DB 430 TCCGTCGTGCAATGCTGAGTGCAGGCGCGCGGATCCGAGAGTCCACCAATCTTGAA 489
QY 162 ProGlyGly-----IleLeuGly-----ValThrGluValValLeuArgGluAlaGly 177
DB 490 CCCGCGCGCGGCTGTCTCTCGAGACATCATCTCGGTTCCGATCCCGGAAAGATGAC 549
QY 178 GlyGlyMetProValSerGlyAspArgTyrProThrGlyLeuArgIleCysLeuAlaGlu 197
DB 550 GCGGCGGCTT-----TGG--ACGGGACACGACCGCCCATCTTGAAC 588
QY 198 GlnLeuLeuGlnSerLeuArgAlaIleGlyPheGluIleLeuAspTyrGluAspValSer 217
DB 589 AGCTTCACGCGCGCTGGTGCAGGACCGGCTTCGAGATTCTCGAATGCACCGACTCTACG 648
QY 218 SerArgThrArgTyrPheMetProGlnPheAlaGlnGluLeu----- 231
DB 649 GCACAGACACAGGTGATGCTCTCTCGTACGTCGACGAGTTCTCCGGAATCTCATAG 708
QY 232 -----AlaAlaHisGlnHis-----GlyIle 238
DB 709 CTCGCCGCGCTGCAGCTGCGGCTGTCGACACTGCAGCAACGCTACTTGGAGACATC 768
QY 239 AlaAspArgTyrGlyProAlaValAlaGlyTrpAlaAlaAlaValCysAspTyrGluLeu 258
DB 769 GCGGCGAGACGCGGCGGAGACACGCTGATGCGCGCGCTTGCGGAAATACCGGAAA 828
QY 259 -----TyrAlaHisAsp-----MetGlyTyrAlaIleLeuThrAlaArgLeu 272
DB 829 CATCCGATTTACGCAAGAAACGAAAGACATGGGTTTCATGCTCTCCAGGCTCGAAG 888

```

```

XX S. spinosa DNA fragment SEQ ID 4.
DE Forsamine; trimethylrhinamose; polyketide synthase; biosynthesis;
XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;
XX macrolide; insecticidal; de.
XX Saccharopolyspora spinosa.
XX DE19957266-A1.
XX 08-MAR-2001.
XX 29-NOV-1999; 99DE-01057268.
XX 27-AUG-1999; 99DE-01040596.
XX (FARB ) BAYER AG.
XX Eberz G, Moehrle V, Froede R, Veltan R, Salas JA;
XX WPI, 2001-267102/28.
XX
XX This invention describes a novel method nucleic acid (I) and its encoded
XX polypeptide (II) containing at least one region that encodes an enzymatic
XX activity involved in biosynthesis of spinosyns. (I) are used (i) to
XX identify, inactive or modulate genes involved in the biosynthesis of
XX (II); (ii) to generate a library of polyketide synthases; (iii) for
XX adding forosamine or trimethylrhinamose to a spinosyn or polyketide
XX aglycone; and (iv) for recombinant production of the corresponding
XX enzymes, which are used for production of (II), their precursors or
XX derivatives, including production of transgenic plants that express (II)
XX and thus have increased resistance to insects. (i) are also useful as
XX markers for sequencing of the Saccharopolyspora spinosa genome. (ii) are
XX macrolides with insecticidal, but not antibacterial, activity, and can
XX also be used to raise specific antibodies, useful for identifying
XX expression clones in a gene bank. Cells transformed with (I) may produce
XX (II) at significantly increased levels or produce new derivatives of
XX (II). This sequence represents a genomic DNA fragment of the S. spinosa
XX genome which contains the coding regions for proteins involved in
XX forosamine and trimethylrhinamose biosynthesis
XX
XX Sequence 45624 BP; 7933 A; 14913 C; 14940 G; 7838 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 3,08e-31 Length: 45624
XX Score: 435.50 Matches: 108
XX Percent Similarity: 51.67% Conservative: 47
XX Best Local Similarity: 36.00% Mismatches: 96
XX Query Match: 30.16% Indels: 49
XX DB: 4 Gaps: 11
XX
XX US-10-069-353a-8 (1-275) x AAF88315 (1-45624)
XX
XX QY 7 ProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThrProLeuLeuAsnSerVal 26
XX DB 7434 CCGACCGCGGATCAGGTGAGTCCATCTTCGATGCGTTG-----GGC 7475
XX
XX QY 27 AlaGlyGlyProCysAlaIleHisIleGlyTyrTrp-----GluAsnAsp 41
XX DB 7476 CACGGGCGTCCC-----CTGCACCACGGTACTGGCGGGCGGGGTATCCGAGAGATGCC 7529
XX
XX QY 42 GlyArgAlaSerTrpGlnGlnIleAlaAspArgLeuThrAspLeuValAlaGluArgThr 61
XX DB 7530 GGTGGCACACCGCTGGTGGATGCTGCCACCACTGACCGACCTGTTCATCACAAGGCC 7589
XX
XX QY 62 ValLeuAspGlyGlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAlaLeu 81
XX DB 7589 GGTGGCACACCGCTGGTGGATGCTGCCACCACTGACCGACCTGTTCATCACAAGGCC 7589

```

[illegible]

XX WP1; 2001-267102/28.

XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for

PT recombinant production of insecticidal spinosyn and their derivatives.

XX

XX

PS- Claim 7; Page 14-31; 354pp; German.

XX

CC This invention describes a novel method nucleic acid (I) and its encoded

CC polypeptide (II) containing at least one region that encodes an enzymatic

CC activity involved in biosynthesis of spinosyns. (I) are used (i) to

CC identify, inactivate or modulate genes involved in the biosynthesis of

CC (II); (ii) to generate a library of polypeptide synthases; (iii) for

CC adding forosamine or trimethylthamose to a spinosyn or polypeptide

CC aglycone; and (iv) for recombinant production of the corresponding

CC derivatives, which are used for production of (II), their precursors or

CC derivatives, including production of transgenic plants that express (II)

CC and thus have increased resistance to insects. (I) are also useful as

CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are

CC macrocides with insecticidal, but not antibacterial, activity, and can

CC also be used to raise specific antibodies, useful for identifying

CC expression clones in a gene bank. Cells transformed with (I) may produce

CC (II) at significantly increased levels or produce new derivatives of

CC (II). This sequence represents a genomic DNA fragment of the *S. spinosa*

CC genome which contains the coding regions for proteins involved in

CC forosamine, trimethylthamose and polyketide synthase biosynthesis

XX

XX

Sequence 50000 BP; 8484 A; 16290 C; 16535 G; 8691 T; 0 U; 0 Other;

SO

Alignment Scores:

Pred. No.: 3,46e-31 Length: 50000

Score: 435.50 Matches: 108

Percent Similarity: 51.67% Conservative: 47

Best Local Similarity: 36.00% Mismatches: 96

Query Match: 30.16% Indels: 49

DB: 4 Gaps: 11

US-10-069-353A-8 (1-275) x AAP88312 (1-50000)

QY 7 ProThiSerGlnGlnValGlyGlnMetTyRAspLeuValThrProLeuLeuAsnSerVal 26

DB 38191 CCGACCGCGGATCAGGAGTGAGTCCATCTTCGATGCGGTG-----GCG 38150

QY 27 AlaGlyGlyProCysAlaIleHisIleGlyTyTrp-----GluAsnAsp 41

DB 38149 CACCGGGCGTCC-----CTGCACCAACGGTTACTGGCGCGCGGTATCGGAGGATGCC 38096

QY 42 GlyArgLaseTTPGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArgThr 61

DB 38095 GGTCCACACCGGTGTGTGATGCTGCGGACCAACTACGACCTGTTCATCCACAAGGCC 38036

QY 62 ValLeuAspGlyGlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAlaLeu 81

DB 38035 GCGGTCCGTCCCGGAGGCGACCTGTTCGACCTGGGCTGGCGGCAATGGGACGCCGATGTC 37976

QY 82 ArgValAlaArgAspAsnAlaIleGlnIleThrGlyTyrIleThrValSerGlnValGlnVal 101

DB 37975 CGTCCGGCATCCGCGACGCGCGCTTCGATGACACCGGAAATCACCTGTAAAGCCCGACATCTC 37916

QY 102 AlaIleAlaAlaAspGlyAlaArgGlyArgGlyLeuSerHisArgValAspPheSerCys 121

DB 37915 GCGCGCGCACCAAGCTTCGCGCAACGACACCGGACTGGCCGCGAGTCTTAAGTTGATCTCA 37856

QY 122 ValAspAlaMetSerLeuProTyRProAspAsnAlaPheAspAlaAlaTTPAlaMetGln 141

DB 37855 GTCCAGCGCGCGCCAGCTGCCCTACCGGAGCGTTCTTCACGCGCGCATGGCGGATGACAG 37796

QY 142 SerLeuLeuGlnMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeuLys 161

DB 37795 TCCGTGTGTCAGATCTGTGACACGAGCGCGCGCATCTCCGAGAGTCCACGAAATCTCGGAA 37736

QY 162 ProGlyGly-----IleLeuGly-----ValThrGluValValLysArgGluAlaGly 177

[illegible]

FT		misc_feature	module 2"
FT			30629. .31621
FT			/tag= aa
FT			/note= "Acyl transferase domain (AT2) : part of extender
FT			module 2"
FT		misc_feature	31697. .32254
FT			/tag= ab
FT			/note= "Dehydratase domain (DH2) : part of extender module
FT			2"
FT		misc_feature	33035. .34072
FT			/tag= ac
FT			/note= "Enoyl reductase domain (ER2) : part of extender
FT			module 2"
FT		misc_feature	34082. .34621
FT			/tag= ad
FT			/note= "Beta-ketoreductase domain (KR2) : part of extender
FT			module 2"
FT		misc_feature	34886. .30295
FT			/tag= ae
FT			/note= "Acyl carrier protein domain (ACP2) : part of
FT			extender module 2"
FT		CDS	35419. .44931
FT			/tag= af
FT			/product= "spnc"
FT			/note= "Spinosyn biosynthesis protein C"
FT		misc_feature	35518. .36786
FT			/tag= ag
FT			/note= "Beta-ketosynthase domain (KS3) : part of extender
FT			module 3"
FT		misc_feature	37108. .38097
FT			/tag= ah
FT			/note= "Acyl transferase domain (AT3) : part of extender
FT			module 3"
FT		CDS	38916. .35374
FT			/tag= Y
FT			/product= "spnc"
FT			/note= "Spinosyn biosynthesis protein C"
FT		misc_feature	38992. .39528
FT			/tag= ai
FT			/note= "Beta-ketoreductase domain (KR3) : part of extender
FT			module 3"
FT		misc_feature	39790. .40035
FT			/tag= aj
FT			/note= "Acyl carrier protein domain (ACP3) : part of
FT			extender module 3"
FT		misc_feature	40102. .41373
FT			/tag= ak
FT			/note= "Beta-ketosynthase domain (KS4) : part of extender
FT			module 4"
FT		misc_feature	41713. .42705
FT			/tag= al
FT			/note= "Acyl transferase domain (AT4) : part of extender
FT			module 4"
FT		misc_feature	43615. .44157
FT			/tag= am
FT			/note= "Beta-ketoreductase domain (KR4) : part of extender
FT			module 4"
FT		misc_feature	44431. .44676
FT			/tag= an
FT			/note= "Acyl carrier protein domain (ACP4) : part of
FT			extender module 4"
FT		CDS	44966. .59752
FT			/tag= ao
FT			/product= "spnd"
FT			/note= "Spinosyn biosynthesis protein D"
FT		misc_feature	45077. .46348
FT			/tag= ap
FT			/note= "Beta-ketosynthase domain (KS5) : part of extender
FT			module 5"
FT		misc_feature	46691. .47674
FT			/tag= aq
FT			/note= "Acyl transferase domain (AT5) : part of extender
FT			module 5"

FT	misc_feature	47753..48310	
FT		/+tag= ar	
FT		/note= "dehydratase domain (DH5) : part of extender module 5"	
FT	misc_feature	49226..49771	
FT		/+tag= ag	
FT		/note= "beta-ketoreductase domain (KR5) : part of extender module 5"	
FT	misc_feature	50009..50254	
Alignment Scores:			
Pred. No.:	6,24e-31	Length:	80161
Score:	435.50	Matches:	108
Percent Similarity:	51.67%	Conservative:	47
Best Local Similarity:	36.00%	Mismatches:	96
Query Match:	30.16%	Indels:	49
DB:	2	Gaps:	11
US-10-069-353A-8 (1-275) x AA221501 (1-80161)			
QY	7	ProThrSerGlnGlnValIglyImeTyrAspLeuValThrProLeuLeuAnsSerVal	26
DB	13562	CCGACCGCGGATCAGGTGAGCGTCATCTTCGATCGCTTG-----GCG	135212
QY	27	AlaGlyGlyProCysAlaIleHisIleGlyTyrTrp-----GluLeuAsp	41
DB	13520	CACGGCGGTGCC-----CTGCACACACGGTTACTGGCGCGCGCGGTATCGGAGATGCC	134678
QY	42	GlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArgThr	61
DB	13466	GGTCCACACCGGTGGTGGATGCTGCGACCACTAACGACTGTTCATCCAGCAAGGC	134070
QY	62	ValLeuAspGlyGlyValAlaArgLeuLeuAspValGlyCysGlyThrGlyGlnProAlaLeu	81
DB	13406	GGCGTCGCTCCCGGAGCGCACCTGTTGCACCTGGCGTGGCGCAATGGCGAGCCGGAATGC	133447
QY	82	ArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGlnVal	101
DB	13346	CGTCCGCGCATGCGCCAGCGCGCTTCGAGTCAACCGGATCAACCGTGAACGCCAGCATCTC	132877
QY	102	AlaIleIleAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSerCys	121
DB	13286	GGCGCGCGCACCGAGCTCGCCACAGACCCGACTGCGCGCATGTTGATTCATCTTA	132222
QY	122	ValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMetGln	141
DB	13226	GTCACGCGCGCCAGCTGCTCCCTACCGGACGGTTCTTTCAGCGCCGACATGGGAGATCGAG	131676
QY	142	SerLeuLeuGlnMetSerGlnProAspArgAlaIleArgGluIleLeuArgValLeuLys	161
DB	13166	TCCGTCGTCGATCTGTCACAGCGCGCGCATCGCGCAGGTCCACCGAATCTCTGGAA	131077
QY	162	ProGlyGly-----IleLeuGly-----ValThrGlnValValLysArgGluAlaGly	177
DB	13106	CCGCGCGCGCGTTCGCTCTCGAGACATCATCATCTGGGTTGACTCCCGGAAGACTAC	130447
QY	178	GlyGlyMetProValSerGlyAspArgTrpProThrGlyLeuArgIleCysLeuAlaGlu	197
DB	13046	GGCGCGGTT-----TGG--ACGGGACAGACCGGCCATACCTTGAAC	130068
QY	198	GlnLeuLeuGlnSerLeuAspArgAlaAlaGlyPheGlnIleLeuAspTrpGluAspValSer	217
DB	13007	AGCTTTCACGCGCGGTCTGACGGAAGCGGGTTCGAGATTCGAAATGTCACCGACTCCAGC	129448
QY	218	SerArgThrArgTyrPheMetProGlnPheAlaGlnGluLeu-----	231
DB	12947	GCACAGACAGTGCATGCTCTCTGGTACGTGCACAGATGCTCCGGAATCTCGATGAG	128868
QY	232	-----AlaAlaHisGlnHis-----GlyIle	238
DB	12887	CTCCGCGCGCGCTGCAGCTCGCGCTGTCGGGACCTACCGAGAAACGCTACTTGGAGAGATCT	128287
QY	239	AlaAspArgTyrGlyProAlaValAlaGlyTyrAlaAlaAlaValCysAspTyrGlyLys	258





QY 226 HisGlyLeuAlaAspArgTyrGlyProAlaValAlaGlyTrrAlaAlaValCysAsp 255  
 DB 2460 CGGGGCTCTCGACGGCGCGCGCGACGAGTGAACGGCGCGACCTGATGAGGC 2519  
 QY 256 TyrGlyLeuTyrAlaHisAspMetGlyTrrAlaIleLeuThrAlaArgLysPro 273  
 DB 2520 TTTCGAGGGGATCCGACATCC--GGTTCGTCCTGATGACCGCGACGGCTCT 2570  
 RESULT 15  
 AAC55802  
 ID AAC55802 strand; DNA; 852 BP.  
 AC AAC55802;  
 XX 19-JAN-2001 (first entry)  
 DT S. Javendulae Mitm encoding DNA sequence.  
 DE  
 XX Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;  
 XX anti-inflammatory; immune-enhancer; immunosuppressant; asthma;  
 XX chronic obstructive pulmonary disease; respiratory inflammation;  
 XX fungicide; pesticide; ds.  
 OS Streptomyces Javendulae.  
 XX  
 PN MO200053737-A2.  
 XX  
 PD 14-SEP-2000.  
 XX  
 PF 10-MAR-2000; 2000MO-US006394.  
 XX  
 PR 12-MAR-1999; 99US-00266965.  
 XX  
 PA (MINU) UNIV MINNESOTA.  
 PA (SHEM) SHERMAN D H.  
 PA (MAOY) MAO Y.  
 PA (VARO) VAROGLU M.  
 PA (HEMM) HE M.  
 PA (SHEL) SHELTON P C.  
 XX  
 PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;  
 DR WPI; 2000-601980/57.  
 DR P-PSDB; AAB32457.  
 XX  
 PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene  
 PT cluster useful for cloning mitomycin biosynthetic genes for elucidating  
 PT the molecular basis of mitosome ring system biosynthesis.  
 PS  
 PS Example 1; Page 262-263; 399pp; English.  
 XX  
 XX This invention relates to isolated and purified nucleic acid molecules  
 CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of  
 CC natural products that contain a variety of functional groups, including  
 CC amino benzozquinone and axiridine ring systems. The S. Javendulae  
 CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes spanning  
 CC 55kb of DNA. The invention includes an expression cassette comprising a  
 CC mitomycin biosynthetic gene operably linked to a promoter, and host cells  
 CC transformed with the cassette. The nucleotide, and protein sequences and  
 CC the transformed host cells of the invention result in antiasthmatic,  
 CC anti-inflammatory, cytostatic, immunomodulatory, and antibiotic  
 CC activities. The nucleotide sequences are used to elucidate the molecular  
 CC basis for the biosynthesis of the mitosome ring system, as well as to  
 CC engineer the biosynthesis of novel natural products, e.g. antibiotics,  
 CC anti-inflammatory agents, anti-cancer agents, immune-enhancers,  
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary  
 CC disease as well as other disease involving respiratory inflammation, or  
 CC cholesterol-lowering agents or as crop protection agents (e.g. fungicides  
 CC or insecticides) as well as biopolymers, e.g., in packaging or biomedical  
 CC applications, or to engineer PHA monomer synthases. Sequences AAC55782-  
 CC C55881, AAC55815-C55849 and AAB32485-832542 represent mitomycin  
 CC biosynthetic gene cluster DNA sequences and encoded proteins. Sequences  
 CC AAC55812-C55814, AAC55850-C55856 and AAC55862-C55869 represent PCR

CC primers used in the cloning of the mitomycin biosynthetic genes  
 XX  
 SQ Sequence 852 BP; 125 A; 316 C; 286 G; 125 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 1,386-27 Length: 852  
 Score: 376.50 Matches: 101  
 Percent Similarity: 50.84% Conservative: 51  
 Best Local Similarity: 33.78% Mismatches: 94  
 Query Match: 26.07% Indels: 53  
 DB: Gaps: 11  
 US-10-069-353a-8 (1-275) x AAC55802 (1-852)  
 QY 2 LeuProGlyGlyAlaProThrSerGlnIleValGlyGlnMetTyrAspLeuValThrPro 21  
 DB 25 CTCGCCATGCCCTCAGCCGCGCTCCGAGAGTGGCGCGCTCTACGACCGGTTCCAGCG 84  
 QY 22 LeuLeuAsnSerValAlaGlyGlyProCysAlaIleIleHisGlyTyrTrpGluAsn-- 40  
 DB 85 CTGGAGCCGCGCTCCCTCGGCGAG-----ACCTGCACTTCGGCTACGGAGATCCCCC 138  
 QY 41 AspGlyArgAlaSerTrrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60  
 DB 139 GACAGCGAGTCCCGCTGGCGCAGGCCACCGGCTCACCGACATGATGCGCGAGCGCG 198  
 QY 61 ThrValLeuAspGlyGlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80  
 DB 199 CTGCGCATCGCGCGCGCTCCGCGCTCCGAGCTTCGCTGCGCTGCGCGTGGAGACCCGGGC 258  
 QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyTyrLeuValSerGlnValGln 100  
 DB 259 GTACGCATGCCCGGCTCAGCGGAGCGCATGTCACCGGAGATCTCGGTGAGCATGAGCAG 318  
 QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120  
 DB 319 GTCTGTCGGGCGCAAGCGCTGGCGCGAGGAGCGCGGCTCGCGACCGCGCGCTCCGAG 378  
 QY 121 CysValAspAlaMetSerLeuProTyrTrpAlaAspAsnAlaPheAspAlaAlaTrpAlaMet 140  
 DB 379 CGGGCCGAGCGCATGAGCTCCCTTCGAGGAGCAGAGAGCTTGACCCCTCATCGCCCTC 438  
 QY 141 GlnSerLeuGlnIleMetSerGluProAspArgAla-----IleArgGlnIleLeuArg 158  
 DB 439 GAATGCATCATCCACATG-----CCGACCGCGCCAGGTGCTGCCACGATCGGCCG 492  
 QY 159 ValLeuLysProGlyGlyIleLeuGlyValThrGlnValValIleArg----- 174  
 DB 493 GTCTCGCGCGCGCGCGCGCTGCTGAGTCAACGACTTCTTCGAGCGGCGCCCTCGCC 552  
 QY 175 --GluAlaGlyGlyGlyMetPro-----ValSer 183  
 DB 553 CCCGAGGGGGCGCGCGCGCTCCAGCTTCCACGACTTCATGATGACCATGATGCTACG 612  
 QY 184 GlyAspArgTrrpThrGlnIleuArg-----IleCys 194  
 DB 613 GCCGAGCGGATACCTCCCTCGCTCGCGGGCGCGCGCTGGCTGAGAGAGATTCCTCGAC 672  
 QY 195 LeuAlaGlnIleuLeuGlnIleuSerLeuArgAlaAlaGlyPheGlnIleLeuAspTrpGlu 214  
 DB 673 ATCAGGAGCAGACACCTCGGAGAGACC-----TTGAGCTGCTCTCG--GAG 717  
 QY 215 AspValSerSerArgThrArgTrrpMetProGlnPheAlaGlnIleuValAlaHis 234  
 DB 718 CGCATCAACTCTCGAAGCAGAGCTGAGACCGCATTCGCGCAGAGATG--GTGAC 774  
 QY 235 GlnHisGlyIleAlaAspArgTrrpGlyProAlaValAlaGlyTrrpAlaAlaValCys 254  
 DB 775 CAGTTGACCCCGCGGACCTCGTGGC----- 801  
 QY 255 AspTrrpGlyLysTrrpAlaHisAspMetGlyTrrpAlaIleLeuThrAlaArgLysPro 273  
 DB 802 -----GTCAAGAGATTCCGGCTATCTGCTGTGCTGCCCGCCAGCCCG 843

Search completed: August 29, 2005, 20:03:36  
Job time : 747 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 29, 2005, 19:36:04 ; Search time 3539 Seconds  
(without alignments)  
2957.806 Million cell updates/sec

Title: US-10-069-353A-8

Perfect score: 1444  
Sequence: 1 VLPGAPTSQQVGMWDLVT.....YKXAHMGVAILTPKPVG 275

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlh  
-O/cn2 1/USPTO.spool/US1006953/runat.26082005.172054.14678/app.query.fasta\_1.455  
-DB=EST -QPM=fastap -SUFFIX=ret -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=numan40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcr -NOR=ext -HEAPSIZE=500 -MINLEN=2000000000  
-USRR=US1006953 @CN2 1.1 3437 @runat.26082005.172054.14678 -NCPU=6 -ICPU=3  
-NO\_MMP -LARGEJURY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:  
1: gb\_ests1:\*  
2: gb\_ests2:\*  
3: gb\_hlc:\*  
4: gb\_ests3:\*  
5: gb\_ests4:\*  
6: gb\_ests5:\*  
7: gb\_ests6:\*  
8: gb\_g881:\*  
9: gb\_g882:\*

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	263.5	18.2	787	6	CA764573 AP53-RpE
2	245.5	17.0	842	7	CN814661 HRO4501 B
3	241	16.7	575	5	B0514932 EST622347
4	241	16.7	859	7	CO076074 GR_EA37D
5	240.5	16.7	857	7	CN008962 WHE2647B
6	240.5	16.7	980	7	CO009336 EST797671
7	240	16.6	824	7	CO129317 GR_Eb260
8	239.5	16.6	902	7	CF515375 CAP0001 I
9	238.5	16.5	669	6	CA220497 SCRUF402

10	238.5	16.5	1665	3	AY103656
11	237.5	16.4	684	7	CA927208
12	236.5	16.4	786	7	CO120890
13	236.5	16.4	961	7	CO009335
14	236	16.3	634	1	AJ431900
15	234	16.2	650	4	Bj271067
16	231.5	16.0	1001	7	CO004512
17	231	16.0	643	5	B0408882
18	230.5	16.0	723	7	CF476020
19	230.5	16.0	755	6	CA064977
20	230.5	16.0	811	4	BM405193
21	230.5	16.0	815	4	BM405276
22	230	15.9	814	7	CN812339
23	230	15.9	854	7	CO092359
24	229	15.9	624	2	AM424526
25	228.5	15.8	592	5	B0990099
26	228	15.8	642	6	CA284871
27	227.5	15.8	592	6	CA101641
28	227.5	15.8	628	2	BF278472
29	227	15.7	600	5	Bf640731
30	226.5	15.7	626	6	CD037447
31	226.5	15.7	842	7	CO086997
32	226.5	15.7	916	7	CO082474
33	226	15.7	861	7	CO117413
34	226	15.7	891	7	CN809395
35	225.5	15.6	688	2	BF479484
36	225	15.6	849	7	CO070289
37	224	15.5	858	7	CK935285
38	222.5	15.4	952	7	CO082835
39	221.5	15.3	1198	3	CN80A23U
40	221	15.3	751	6	CB289364
41	219.5	15.2	629	6	CA170495
42	218.5	15.1	771	7	CO367476
43	218	15.1	544	2	BF598669
44	218	15.1	587	4	BG363118
45	217	15.0	570	2	AW706079

## ALIGNMENTS

RESULT 1  
CA764573  
LOCUS  
DEFINITION  
CA764573  
AP53-RpE 04 B21 T7 081 ab1 IRRI drought stress Panicle Library  
Oryza sativa (indica cultivar-group) cDNA clone C0001197 5' similar  
to Delta (24)-sterol C-methyltransferase (EC 2.1.1.41), mRNA  
sequence.

ACCESSION  
CA764573  
VERSION  
CA764573.2 GI:27546528  
KEYWORDS  
EST.  
SOURCE  
Oryza sativa (indica cultivar-group)  
ORGANISM  
Oryza sativa (indica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE  
AUTHORS  
Bennett,J., Arumugam,K., Lafitte,R., Wen,J., Rudd,S. and  
Bruskiewich,R.M.  
IRRI Drought Stress Panicle cDNA Library  
Unpublished (2002)  
COMMENT  
On Dec 2, 2002 this sequence version replaced gi:25993828.  
Contact: Richard Bruskiewich  
Biometrics and Bioinformatics Unit  
International Rice Research Institute  
DAPO 7777, Metro Manila, Philippines  
Tel: +63-2-845-0563  
Fax: +63-2-845-0606  
Email: r.bruskiewich@cgiar.org  
International Rice Information System (IRIS):  
http://www.iris.irri.org: D0201196  
Assignment of putative function to the sequence by S. Rudd of the  
Munich Information Center for Protein Sequences  
(http://mips.gsf.de)



```
Db      128 GAATCTTTTCACTTGTCTAC---AGTGGAAATGGGAA-----TCTTTAAGGAA 175
Qy      49 AlaAlaSpArgLeuThrAspLeuValAlaGluArgThrValLeuAspGlyValArg 68
Db      176 AGATCAAAACGACACGAGCATTTTGGCCCTTCAGCTTGAGTGAACACGAGGATGAAG 235
Qy      69 LeuLeuAspValGlyCysGlyThrGlyGlnProAlaLeuArgValAlaArgAspAla 88
Db      236 GTTTTGACCTGGCTGGTGGAAATGGTGAACCTTAAGAGATTCACGATTACTCTCA 295
Qy      89 IleGlnIleThrGlyIleThrValSerGlnValAlaAlaIleAlaIleAspCysAla 108
Db      296 ACCCTGATTACCGGATTGAACAACAACGACTATCAATCACTGAGGGAAGAGCTCAAT 355
Qy      109 ArgGluArgGlyLeuSerHisArgValAspPheSerCysValAlaMetSerLeuPro 128
Db      356 CGTCTGGCAGGACTTGTGGAATGGAATGTTGTCAGGACAGACTTCATGAAATATGCTT 415
Qy      129 TyrProAspAsnAlaPheAspAlaAlaTPrAlaMetGlnSerLeuGluMetSerGlu 148
Db      416 TTCTCCGATTAACACTTTTGATGCTGTTATGATGAGCAACATGTCATGACCTGAT 475
Qy      149 ProAspArgAlaIleArgGluIleLeuArgValLeuLysProGlyGlyIleLeuGlyVal 168
Db      476 CCGGTTGGCTGCTATTAAGAGATCTAACCGGTATTAACCTGGGACGCTTTTCTGCTGA 535
Qy      169 ThrGlu-----ValValLysArg----- 174
Db      536 TATGAGTGGCTACTTACTGATCACTATGATCCAAACAATGCCAACCAAGAGATTAAG 595
Qy      175 -----GluAlaGlyGlyGlyMetProValSerGlyAspArgTyrProThrGlyLeu 191
Db      596 GATGAATATGAGCTGGTATGATGCTACCA----- 625
Qy      192 ArgIleCysLeuAlaGluGlnLeuLeuGlnSerLeuArgAlaIleArgPheGluIleLeu 211
Db      626 GATATCAAGAACCACTGACCAATGCTTCAAGCTGTTAAAGATGAGGATTTAGAGTTAT 685
Qy      212 AspTyrGluAsp-----ValSerSerArgThrArgTyrPheMetPro 225
Db      686 ----TGGGATTAAGATCTTGCTGTAGATTCTCCATGCTTGATCTTGATCTTGCC 733

RESULT 3
BO514932      575 bp      mRNA      linear      EST 07-MAR-2003
LOCUS      BO514932
DEFINITION      EST622347 Generation of a set of potato cDNA clones for microarray
analyses mixed potato tissues Solanum tuberosum cDNA clone STM1P11
5' end, mRNA sequence.
ACCESSION      BO514932
VERSION      BO514932.1 GI:21373801
KEYWORDS      EST.
SOURCE      Solanum tuberosum (potato)
ORGANISM      Solanum tuberosum
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum.
1 (bases 1 to 575)
Buell, C.R., Hart, A., Baker, B., Tanksey, S., Fry, W., Smart, C.,
Reestepo, S., Griffiths, H., van der Hoeven, R., Tesi, J. and
Karimicheva, S.A.
Generation of a set of potato cDNA clones for microarray analyses
Unpublished (2002)
Other ESTs: EST622348
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: T3.
Location/Qualifiers
```

```
..      source
..      1..575
..      /organism="Solanum tuberosum"
..      /mol_type="RNA"
..      /cultivar="Kennebec or Binje"
..      /db_xref="taxon:4113"
..      /clone="STM1P11"
..      /tissue_type="mixed tissues"
..      /lab_host="SOLR"
..      /clone_lib="Generation of a set of potato cDNA clones for
..      microarray analyses mixed potato tissues"
..      /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
..      XhoI; supplier: Combination of untreated and Phytophthora
..      infestans-created libraries of stolons, leaves, leaflets,
..      axillary buds of stem explants, petioles, germinating
..      eyes, tubers, or roots."

ORIGIN
Alignment Scores:
Pred. No.:      9.52e-16      Length:      575
Score:      241.00      Matches:      64
Percent Similarity:      48.73%      Conservative:      32
Best Local Similarity:      32.49%      Mismatches:      72
Query Match:      16.69%      Indels:      29
DB:      5      Gaps:      4

US-10-069-353A-8 (1-275) x BO514932 (1-575)
Qy      45 SerTyrGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArgThrValLeuAsp 64
Db      7 TCACCTCAGAGAAACGATTAAGAGCATGACGACTTCTTGCCATTCGAATTGGGATTAAG 66
Qy      65 GlyGlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAlaLeuArgValAla 84
Db      67 CCAAGCAAAAGCTCTTGATGTAGAGATGGAATTGGAGGCCCTTAAGACAAATGCT 126
Qy      85 ArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValAlaIleAla 104
Db      127 CGATTACGCTTCACTACGATTACAGGCTTCAACAACAATGATATCAGATCTAGCGGA 186
Qy      105 AlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSerCysValAspAla 124
Db      187 CAGGTGTGAACCGCAAGTGGGATGGATTCACAACTTGCACTTGTAAGGGATGATTC 246
Qy      125 MetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTPrAlaMetGlnSerLeuLeu 144
Db      247 ATGAATAATCCATTTCCCTGACATAGCTTGTATGACGATTAAGCAATTAAGACTTACCTGC 306
Qy      145 GluMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeuLysProGlyGly 164
Db      307 CATGACACCGATCCGTTGGATGCTATAGAGATCTATAGGCTGTGAAGCTGGTCAAG 366
Qy      165 IleLeuGlyVal-----ThrGluValVal 172
Db      367 TGGTTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 426
Qy      173 LysArg-----GluAlaGlyGlyGlyGlyMetProValSerGlyAspArgTyr 187
Db      427 AAAAGATCAAGAGAAATTAAGTGCCTGGAATATGCTTCG----- 468
Qy      188 ProThrGlyLeuArgGlyCysLeuAlaGluGlnLeuLeuGlnSerLeuArgAlaIleArg 207
Db      469 -----GAGATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 516
Qy      208 PheGluIleLeuAspTyrGluAsp-ValSerSerArgThrArgTyrPhe 223
Db      517 TTTGAAGTTGTG---TGGGATTAAGATTCGCTGAAGACTCACTCGTTT 562

RESULT 4
CO076074      859 bp      mRNA      linear      EST 15-JUN-2004
LOCUS      CO076074
DEFINITION      GR_Ea37D20.r GR_Ea Gossypium raimondii cDNA clone GR_Ea37D20 3',
mRNA sequence.
ACCESSION      CO076074
Location/Qualifiers
```



bluescript SK, Site\_1: EcoRI; Site\_2: XhoI; Plants were grown in the greenhouse. Spikes were sprayed at anthesis with Fusarium graminearum. Total RNA, and poly(A) RNA were prepared and pooled from infected spike at 0, 6, 12, 24, 36 and 48 hours after inoculation, a cDNA library was made, and the cDNA clones were in vivo excised to give Bluescript phagemids in G. Muenbauer lab at the University of Minnesota (Kruger, W.M., Muenbauer, G.J., Pritsch, C., Vance, C.). The cDNA library should contain genes of both wheat and fungal pathogen origin. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

## ORIGIN

## Alignment Scores:

Pred. No.:	1,11e-15	Length:	587
Score:	240.50	Matches:	57
Percent Similarity:	49.21%	Conservative:	36
Best Local Similarity:	30.16%	Mismatches:	69
Query Match:	16.66%	Indels:	27
DB:	7	Gaps:	2

US-10-069-353a-8 (1-275) x C0008962 (1-587)

```

Qy      48 GlnAlaAlaAspArgLeuThrAspLeuValAlaGluArgThrValLeuAspGlyGlyVal 67
Db      23 CAGGCCATTGCCCGTCATGACGACTACCTTCCACACAGATCGGTATCAAGAGCGGATG 82
Qy      68 ArgLeuLeuAspArgValGlyCysGlyThrGlnProAlaLeuArgValAlaArgAspAsn 87
Db      83 AAGGTTCTTATGTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 142
Qy      88 AlaIleGlnIleThrGlyIleThrValSerGlnValAlaAlaAlaAlaAspCys 107
Db      143 GGTGCTCATATTTCACCGCCCTGAACACAACTACAGATCGAGCTACCCACTAC 202
Qy      108 AlaArgGluArgGlyLeuSerHisArgValAspPheSerCysValAlaAlaMetSerLeu 127
Db      203 GCCTTCAGAGAGGGTGTCTCCACACGACTGAGTTCTCAAGGGTGAATTCATGCAATG 262
Qy      128 ProTyrrProAspAsnAlaPheAspAlaAlaThrAlaMetGlnSerLeuLeuGluMetSer 147
Db      263 TCTTTCCTCCCAACACGCTTGTGATGCGTTATCCATAGGACCCGCTTACGCTCC 322
Qy      148 GluProAspArgAlaIleArgGluIleLeuArgValLeuArgProGlyGlyIleLeuGly 167
Db      323 ACTCTCAAGGCGATCTACAGGAGATTTCCGTCTTAAAGCCGGTGGTCTTGGT 382
Qy      168 ValThrGluValValValArg----- 174
Db      383 GTGACGAGTGGCTCATGACTGACGAGTACGACAACTCCGCCACCGTAGATT 442
Qy      175 -----GluAlaGlyGlyGlyMetProValSerGlyAspArgTrrProThrGly 190
Db      443 CGATTGGGATCGAACAGGGTGAATGTAATTC----- 475
Qy      191 LeuArgIleCysLeuAlaGluGlnLeuLeuGlnSerLeuArgAlaAlaGlyPheGluIle 210
Db      476 ---AACATGTCCAGAGCTCTGAGGGATCCCTCCATCCAGACTCGGTTGGAATG 532
Qy      211 LeuAspTrrGluAspValSerSerArg 219
Db      533 CTCACACACGAGATCTTGGCGATCGC 559

```

## RESULT 6

C0009336

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

980 bp mRNA linear EST 09-JUN-2004  
 EST797671 Coccidioides posadaei spherule cDNA library, 0.4 to 2.3  
 kb Coccidioides posadaei cDNA clone CIBB26 5' end, mRNA sequence.

C0009336

EST

Coccidioides posadaei

## ORGANISM

Coccidioides posadaei  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 Onygenales; mitosporic Onygenales; Coccidioides.

## REFERENCE

1 (bases 1 to 980)

Gardner/M.J. and Cole,G.T.

Analysis of gene expression in Coccidioides posadaei mycelia and

spherules via expressed sequence tags

Unpublished (2003)

Other\_ESTs: EST797670

Contact: Gardner MJ

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301 838 3519

Fax: 301 838 0208

Email: gardner@igr.org

Seq primer: M13 Reverse.

Location/Qualifiers

1..980

/organism="Coccidioides posadaei"

/mol\_type="mRNA"

/strain="C735"

/db\_xref="taxon:199306"

/clone="CIBB26"

/dev\_stage="spherules"

/lab\_host="E. coli DH10B, T1 phage resistant"

/clone\_1lb="Coccidioides posadaei spherule cDNA library, 0.4 to 2.3 kb"

/note="Vector: pExpress 1; Site\_1: Not I; Site\_2: Eco RV; Coccidioides posadaei spherule cDNA library, 0.4 to 2.3 kb"

## ORIGIN

## Alignment Scores:

Pred. No.:	2.24e-15	Length:	980
Score:	240.50	Matches:	67
Percent Similarity:	47.93%	Conservative:	37
Best Local Similarity:	16.66%	Mismatches:	90
Query Match:	7	Indels:	23
DB:	7	Gaps:	5

US-10-069-353a-8 (1-275) x C0009336 (1-980)

```

Qy      16 TyrAspLeuValThrProLeuLeuAsnSerValAlaGlyPro-----CysAla 32
Db      12 TACAACTCGCGAGGACATGTCGAATATGATGGGCGGCTTCTTCCACTTGTGCT 71
Qy      33 IleHisGlyTyrTrrGluAsnAspGlyArgAlaSerTrrGlnGlnAlaAlaAspArg 52
Db      72 TTCCCGATATGC-----GAACCTTCCGTCAGGCTATGCTCGA 110
Qy      53 LeuThrAspLeuValAlaGluArgThrValLeuAspGlyValArgLeuLeuAspVal 72
Db      111 CAGAACATATATCTCGACACTCCATCGCGCTCCAGAGGGCCAGATGTTGATGATC 170
Qy      73 GlyCysGlyThrGlyGlnProAlaLeuArgValAlaArgAspAlaIleGlnIleThr 92
Db      171 GGCTGCGCGCTTGCGGCGCCCGCAAGAAATTCCTCCGCGGTCAATATTGTT 230
Qy      93 GlyIleThrValSerGlnValAlaGlnAlaIleAlaAlaAspCysAlaArgGluArgGly 112
Db      231 GGTTAATAATACACAGCATCTCAAAATTGAACGGGCACTCGATATGCGGGAAGAGT 290
Qy      113 LeuSerHisArgValAspPheSerCysValAspAlaMetSerLeuProTyrrProAspAsn 132
Db      291 TTGTCAAGCAGCTCCGCTTACCAAGGGTGAATTTATGCAATGCTCTTGAACCAAT 350
Qy      133 AlaPheAspAlaAlaThrAlaMetGlnSerLeuLeuGlnMetSerGluProAspArgAla 152
Db      351 ACATTTATCTGTGATATGCAATTGAGGCTACAGCTTACGCTCTTCCGTTAGAGGGCTC 410
Qy      153 IleArgGluIleLeuArgValLeuArgProGlyGlyIleLeuGlyValThrGluValAla 172
Db      411 TACAGACAGATTTCATTAACCTATTAAACCCGGCGGTGTTTGGTGTATACGAATGGTCT 470

```

QY 173 LysargLysAlaA---GlyGlyGlyMetProVal-----182  
Db 471 ATGACCGACGCTTATGATACAAACAACCCCGTCACCGTGAATCAGACTGGTATCGAA 530  
QY 183 SerGlyAspArgTrpProThrGlyLeuAglGlyLeuAglGlnLeuLeuGlnSer 202  
Db 531 CAAAGGAAATGGAATTTCCAAACATGATGTCAG-----GCCAACCAAGCCCTTGATGCC 581  
QY 203 LeuArgAlaAlaGlyPheGlnIleLeuAspTrpGluAspValSerSerArg 219  
Db 582 TTTAAGCCCGCCCGTTTCGAACTCAATCAAGCTGAGAGATCTAGACACCT 632  
RESULT 7  
LOCUS COI29317 824 bp mRNA linear EST 17-JUN-2004  
DEFINITION GR\_Eb26017.r GR\_Eb Gossypium raimondii cDNA clone GR\_Eb26017.3',  
mRNA sequence.  
ACCESSION COI29317 GI:48878295  
VERSION COI29317.1 GI:48878295  
KEYWORDS EST.  
SOURCE Gossypium raimondii  
ORGANISM Gossypium raimondii  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eustosids II; Malvales; Malvaceae; Malvoideae; Gossypium.  
1 (bases 1 to 824)  
Kim,H., Yu,Y., Kudrna,D., Hatfield,J., Stum,D., Mueller,C.,  
Udall,J.A., Rapp,R.A., Wendel,J.F., Rao,K., Soderlund,C. and  
Wing,R.A.,  
TITLE Global assembly of Cotton ESTs  
JOURNAL Unpublished (2004)  
COMMENT Contact: Rod A. Wing  
Arizona Genomics Institute  
The University of Arizona  
Forbes Building Room 303, Tucson, AZ, 85721-0036, USA  
Tel: 520 626 9595  
Fax: 520 621 1259  
Email: http://genome.arizona.edu  
Plate: 26 row: O column: 17.  
FEATURES  
source  
1..824  
Location/Qualifiers  
/organism="Gossypium raimondii"  
/mol\_type="mRNA"  
/db\_xref="taxon:29730"  
/clone="GR\_Eb26017"  
/tissue\_type="floral"  
/dev\_stage="3 to +3 DPA"  
/lab\_host="DH10B"  
/clone\_lib="GR\_Eb"  
/note="Vector: PCMV.SPORT-6.1; Site 1: NotI; Site 2:  
ECORV; Library made by Invitrogen with RNA supplied by  
Wendle lab. Directional cloned into NotI-EV. Clones  
plated/picked by AGI. More glycerol clones held in -80."

## ORIGIN

## Alignment Scores:

Pred. No.: 2e-15 Length: 824  
Score: 240.00 Matches: 69  
Percent Similarity: 43.15% Conservative: 38  
Best Local Similarity: 27.82% Mismatches: 107  
Query Match: 16.62% Indels: 34  
Db: 7 Gaps: 8

US-10-069-353A-8 (1-275) x COI29317 (1-824)

QY 3 ProGlyGlyAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThrProLeu 22  
Db 35 CCCGATGCACGTGATTCGACGAGGATGCGAGTTCTACGACCACTGCTGGGCTT 94  
QY 23 LeuAnserValAlaGlyGlyProCysAlaIleHisGlyTyrTrpGluAsnAspGly 42  
Db 95 TGGGAAGACATTTGGGGTGAC-----CATATGACATCATGTTTATACGATCGGATTC 148

QY 43 ArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArgThrVal 62  
Db 149 AATGTTTACAGGCTCCGATCATCATCAGCTGCCAGATCCGAATGATC---GAAGAAATCGCTC 205  
QY 63 LeuAspGlyGlyVal-----ArgLeuLeuAspValGly 73  
Db 206 CGTTTCGCGTGAATATCCGATATCCAGATCCAGCAAGCCCAAGCAATGATGATTTGGA 265  
QY 74 CysGlyThrGlyGlnProAlaLeuArgValAlaArgAspAsnAlaIleGlnIleThrGly 93  
Db 266 TTTGGGATRAGAGGACGCTCTGATCATCTTGAAGAAATTTGGACGCAATGTCAGGC 325  
QY 94 IlerValSerGlnValGlnValAlaIleAlaAspCysAlaArgGluArgGlyLeu 113  
Db 326 ATTACTTTGAGCCCTCGTCCAGCTCGAAGGGCTATGTTCTTGAAGATGAAGACGTA 385  
QY 114 SerHisArgValAspPheSerCysValAspAlaMetSerLeuProTyrPheAspAsnAla 133  
Db 386 GCAGACAAGGTTTCATTTCAAGTTGCAGATGCTCTGAAGCAACCATTCCTGATGTCAG 445  
QY 134 PheAspAlaAlaTrpAlaMetGlnSerLeuGlnMetSerGluProAspArgAlaIle 153  
Db 446 TTTGATCTAGTTGGTCTATGGAAGAGTGAGAACACATGCTGATPAAAGCTTAAGTTGTT 505  
QY 154 ArgGlnIleLeuArgValLeuAspProGlyGlyIleLeuGlyValThrGluValVallys 173  
Db 506 AATGAGTTGGACGAGGTTGACGCTCTCGAGGCACATATATATGATGACATGCGCCAT 565  
QY 174 ArgGluAlaGlyGlyGlyMetProValSerGlyAsp-----ArgTrpProThrGlyLeu 191  
Db 566 AAGGATCTTGGT-----CCCTCAGAGAGGATTTGAGCCCATGGAGAAAAAGCTG 616  
QY 192 -----ArgIleCysLeuAla-----Glu 197  
Db 617 CTAAATAGAGATATGTAATGCTTACTTACTGATGCTGTCTTACTTCTGACTATGTC 676  
QY 198 GlnLeuLeuGlnSerLeuArgAlaAlaGlyPheGlnIleLeuAspTrpGluAspValSer 217  
Db 677 AAACACTTTCAGTCCCTATCTCTCCAGATATPAAAGCTGCTGATTTG-----TCT 727  
QY 218 SerArgThrArgTyrPheMetPro 225  
Db 728 GAGCGTGTGCCCGCTTTGGCCG 751  
RESULT 8  
LOCUS CF515375 902 bp mRNA linear EST 09-SEP-2003  
DEFINITION CAP0001\_IVF\_c10 Vitis vinifera cv. cabernet sauvignon (Clone 8)  
petiole - CAP Vitis vinifera cDNA clone CAP0001\_IVF\_c10 5', mRNA  
sequence.  
ACCESSION CF515375  
VERSION CF515375.1 GI:34547143  
KEYWORDS EST.  
SOURCE Vitis vinifera  
ORGANISM Vitis vinifera  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; Vitaceae; Vitis.  
1 (bases 1 to 902)  
Goed da Silva,F., Iandolo,A., Lim,H., Baek,J., Leslie,A., Xu,J.,  
Jones,K. and Cook,D.  
Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon'  
berries at various developmental stages  
Unpublished (2003)  
CONTACT: Douglas Cook, PhD  
CAES Genome Facility  
UC Davis, Plant Pathology  
One Shields Ave, Davis, CA 95616, USA  
Tel: 530 754 6561  
Fax: 530 754 6617  
Email: drcoc@ucdavis.edu  
Seq primer: ACGTACCGACATATGCGC.



# FEATURES

Location/Qualifiers  
 1. 902  
 /organism="Vitis vinifera"  
 /mol\_type="mRNA"  
 /cultivar="Cabernet Sauvignon"  
 /db\_xref="taxon:29760"  
 /clone="CAP0001 IVF C10"  
 /sex="Hermaphrodite"  
 /dev\_stage="Onset of veraison (berry softening)"  
 /lab\_host="Vitis vinifera cv. cabernet sauvignon (clone 8) Petiole - CAP"  
 /note="Organ: Petiole; Vector: pDNR; Site 1: SfiI; Site 2: SfiI; CAP is a cDNA library of Vitis vinifera cv. 'Cabernet Sauvignon' clone 8 petioles. Samples were collected on July 10, 2002 from plants on the onset of veraison (berry softening). Sampled vines were located at the University of California, Davis, Experimental vineyard. cDNAs were made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:  
 5'-AACGAGTGTATCAGCGAGGCGGCGGCGG-3' and  
 5'-ATTCTAGAGCGGCGGCGGCGGCGGCGG-3'. Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 Kb size fraction."

## ORIGIN

### Alignment Scores:

Pred. No.: 2,57e-15 Length: 902  
 Score: 239.50 Matches: 81  
 Percent Similarity: 39.86% Conservative: 37  
 Best Local Similarity: 27.36% Mismatches: 89  
 Query Match: 16.59% Indels: 89  
 DB: 7 Gaps: 12

US-10-069-353a-8 (1-275) x CFS15375 (1-902)

```

QY 33 ILEHIEHIEGLYTRTPGUAHNPGLYARGHASERTTPGLINGIAIAIAAPARG 52
   ::::::::::::::::::::
DB 13 ATGCACACACGGCTTCTACGACCGGACTCCGCGCTTCA-----GATGCCGACCAACCGG 66
   ::::::::::::::::::::
QY 53 LEUTHRASPLEU---VALAIAAGIARGTHRAILEUASPLGLYVALA----- 67
   ::::::::::::::::::::
DB 67 TTGGCTCAGATCCGATGATGAGAGAGAGCTCCGGTCCCGAGTTTCCGAGAGGGA 126
   ::::::::::::::::::::
QY 68 -----ARGLEUASPLVALIGLYCYSEGLYTHRGILYINPROALAEUARG 82
   ::::::::::::::::::::
DB 127 GAGAAAGGCCCAAGAGAGAGTGGGATGGGGGTGGGATGGAGGAGACTCAAGGTAC 186
   ::::::::::::::::::::
QY 83 VALAIAARGHAPENALALEGINILETHRGILYIETHRVASERGINVALAIA 102
   ::::::::::::::::::::
DB 187 TTGGCTAAGAGATGAGGCGCAGTTGCCAAGGGATTTACTCCCTCCCAAGCTCA 246
   ::::::::::::::::::::
QY 103 ILEAIAIAAPRYCSAIAARGIARGIYLEUSERTHARSARGVALASPHESERCYSVAL 122
   ::::::::::::::::::::
DB 247 AGGCTCAGACTCTTGCTGCATCCCAAGGTTGGCTACACAGGATTTTCCAAAGTTGCA 306
   ::::::::::::::::::::
QY 123 ASPALWASERLEUPROYTYRPROASPAANALAPHEASPAIAIAATRALAMETGINSER 142
   ::::::::::::::::::::
DB 307 GATGCTCTGATCAACCGTTTCCGATGACAAATTTATCTGTTGGTCCATGAGAGAGT 366
   ::::::::::::::::::::
QY 143 LEUENGLUWESERGIUAPROASPARGALILEARGIUILEUARGVALLEUYSFRO 162
   ::::::::::::::::::::
DB 367 GGAGAACACATGCTGACAAATAAAAGTTTGTAGTAGTGGCTCGAGTTGACGCCCA 426
   ::::::::::::::::::::
QY 163 GLYGLY---ILEUENGLYVALTHR-----GLUVAL 171
   ::::::::::::::::::::
DB 427 GGTGGCAANTAAATCTTGTAACATGGTGCCATAGGAGATCTTCCCTTGAAGAAATCC 486
   ::::::::::::::::::::
QY 172 VALIYASRGIAUAGIAGIYGLIYMETPROVALSERGIYASPARGTTPROTHRGILYEU 191
   ::::::::::::::::::::
DB 487 TTAAAGCCAGAGGGAAGACATCTTGAC----- 516

```

```

QY 192 ARGILECYSLEUAI-----GLUGLILEU 199
   ::::::::::::::::::::
DB 517 AGATTTCAGAGCTTATTATCTCCAGATGGGTCTTACACTGATTATGTCATAATTA 576
   ::::::::::::::::::::
QY 200 LEUGLUSERLEUARGALAIAGIYPHEGLUILEUASPTTPGLUASPVASERARG 219
   ::::::::::::::::::::
DB 577 CTGAGTCTCTATCTTCTTACAGATATCAAGCGCGGAGATGGTCTGAG----- 624
   ::::::::::::::::::::
QY 220 THRARGTYRPHMECPROGLINPHEALAGIUILEUAIAGLHIEGLIYLEAIA 239
   ::::::::::::::::::::
DB 625 -----TATGTTGACACC----- 636
   ::::::::::::::::::::
QY 240 ASPARGYRGLYPROALAVAI----- 246
   ::::::::::::::::::::
DB 637 -----TTTGGCCACGAGTGAATAGCTGTGCTTTCAGCTTTCAGGGCTTATCACTG 690
   ::::::::::::::::::::
QY 247 -----ALAGIYTP-----ALAIALAVAIYASPTYRGLIYSTRYALA 260
   ::::::::::::::::::::
DB 691 TTACGACAGTGAAGAAACATTAAGAGAGCATGTGGATGCCATTGATGATCCGAGG 750
   ::::::::::::::::::::
QY 261 HISAPMETGLY-----TYRALALEUENHRAIARGIYSPRO 273
   ::::::::::::::::::::
DB 751 TTAAGAGATGGCGCTGATTAAATTGCTATCATTAATGTCGAAGCCC 798
   ::::::::::::::::::::

```

## RESULT 9

CA220497

LOCUS 669 bp mRNA linear EST 25-SEP-2003  
 DEFINITION SCRFL4023B12.g FL4 Saccharum officinarum cDNA clone SCRFL4023B12

ACCESSION CA220497

VERSION CA220497.1 GI:35273861

KEYWORDS EST

SOURCE Saccharum officinarum

ORGANISM Saccharum officinarum

REFERENCE 1 (bases 1 to 669)  
 Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.  
 The libraries that made SUCEST  
 Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

JOURNAL  
 COMMENT Centro de Biologia Molecular e Engenharia Genetica  
 Universidade Estadual de Campinas  
 Caixa Postal 6010, 13083-970, Campinas SP, Brazil  
 Tel: 55 19 3788 1137  
 Fax: 55 19 3788 1089  
 Email: parruda@unicamp.br

Clone distribution: clone distribution information can be found  
 through the Brazilian Clone Collection Center (BCCC) at  
 http://www.bcccenter.fcav.unesp.br  
 Plate: 023 row: B column: 12  
 Seq primer: T7 Promoter Primer.

## FEATURES

source

1. 669  
 /organism="Saccharum officinarum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4547"  
 /clone="SCRFL4023B12"  
 /lab\_host="DH10B"  
 /clone\_1ib="FL4"  
 /note="Organ: Developed inflorescence and rachis (20cm-long); Vector: pSport1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from [developed inflorescence and rachis (20cm-long)]. cDNA was prepared from polyA+ mRNA using Superscript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a Sephadose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at  
 http://succest.lad.ic.unicamp.br/public"

## ORIGIN

## Alignment Scores:

Pred. No.:	2,19e-15	Length:	669
Percent Similarity:	238.50	Matches:	65
Best Local Similarity:	46.02%	Conservative:	39
Query Match:	28.76%	Mismatches:	91
	16.52%	Indels:	31
		Gaps:	5

US-10-069-353a-8 (1-275) x CA220497 (1-669)

```

QY 9 SerGInGInValGlyGInMetYrAspLeuValThrProLeuLeuAsnSerValAlaGly 28
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 22 ACTGATATGTTAAACAATATATATATATGATCTTGCCACTGACCTGCTGATGATGCGGGGT 81
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 29 GlyProCysAlaIleHisGlyYrTrpGluAsnAspGlyArgAlaSerTrpGlnGln 48
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 82 GAATCCTTCACCTTGGCTCAC---AGATGGAATGGAGAA-----TCCTTAGCTGAA 129
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 49 AlaAlaAspArgLeuThrAspLeuValAlaGluArgThrValIleuAspGlyValAlaArg 68
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 130 AGCATCAACCGTCATGAGCATTTTCTTGCCCTGCAGCTTGCGCTTGAAACCGAATGAAAG 189
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 69 LeuLeuAspValGlyCysGlyThrGlyGlnProAlaLeuArgValAlaArgAspAsnAla 88
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 190 GTTTTGATGTGGGCTGCGGAATGAGTGAGCACCTGAGAAATTCGAAAGATTAGCTCA 249
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 89 IlegInIleThrGlyIleThrValSerGlnValGlnValAlaIleAlaAlaAspCysAla 108
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 250 ACTTCAGTTACTGGGTGGAATTAACAATGAAATCCAGATTAACGAGGGGAAAGAGCTCAAT 309
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 109 ArgGluArgGlyLeuSerHisArgValAspPheSerCysValAspAlaMetSerLeuPro 128
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 310 CGGTGGCGAGGAGTTAGTGAACCTGCCATTTTGTTCAGACGACATTTATGAAATGCCG 369
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 129 TyrProAspAsnAlaPheAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 148
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 370 TTTGATGATTAACACTTTTGTATGCTGTTTATGCAATTGAGGCGACATGTATCATCAGCTGAT 429
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 149 ProAspArgAlaIleArgGluIleLeuArgValIleuValLeuValProGlyGlyIleLeuGlyVal 168
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 430 CCAAGTTGGCTGTATTAAGAGATATACCGTGTGTGAACCTGGCGCAGTCTTCCTCTGA 489
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 169 ThrGlu-----ValIleValArg-----ValIleValArg----- 174
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 490 TATGAGTGTGCTTATCTATCATCATATGATCCAAACACGCAACCATTAAGAGATTAAG 549
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 175 -----GluAlaGlyGlyGlyMetProValSerGlyAspArgTrpProThrGlyLeu 191
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 550 GATGAATATTGAGCTTGGCAATGGCGCTGCCA----- 579
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 192 ArgGlyLeuLeuAlaGluGlnLeuLeuGlnSerLeuValArgAlaGlyPheGluIleLeu 211
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 580 GATATCGAAGACACTAGGCAATGTCTTCAGGACGTCGAAGATCGGTTTGTAGGTGTT 639
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 212 AspTrpGluAspValSer 217
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 640 GGAGATTAAGATCTAGCT 657
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

## RESULT 10

LOCUS	AY103656	1665 bp	mRNA	linear	HTC	16-OCT-2002
DEFINITION	Zea mays PC0072714 mRNA sequence.					
ACCESSION	AY103656					
VERSION	AY103656.1	GI:21206734				
KEYWORDS	HTC.					
SOURCE	Zea mays					
ORGANISM	Zea mays					
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.					
REFERENCE	1 (bases 1 to 1665)					

## AUTHORS

## TITLE

Hainey, C.F., DoJan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.

## JOURNAL

Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes

## REFERENCE

## AUTHORS

Unpublished (2002)  
2 (bases 1 to 1665)  
Coe, E.H.

## TITLE

Direct Submission  
Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA

## COMMENT

If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSU, [maizemap.org](http://maizemap.org); ZmDB, [www.zmdb.iastate.edu](http://www.zmdb.iastate.edu); TIGR, [www.tigr.org](http://www.tigr.org); or NCBI, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov). When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: [www.zmdb.iastate.edu](http://www.zmdb.iastate.edu).

## FEATURES

## source

Location/Qualifiers  
1..1665  
/organism="Zea mays"  
/mol\_type="mRNA"  
/db\_xref="MaizeDB:633975"  
/db\_xref="taxon:4577"  
/clone\_lib="Maize Mapping Project/Dupont Consensus Library"

/note="this sequence is part of a project of EST assemblies resulting from the application of public configs to seed Dupont configs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

## ORIGIN

## Alignment Scores:

Pred. No.:	7.64e-15	Length:	1665
Score:	238.50	Matches:	67
Percent Similarity:	46.19%	Conservative:	42
Best Local Similarity:	28.39%	Mismatches:	94
Query Match:	16.52%	Indels:	33
		Gaps:	6

US-10-069-353a-8 (1-275) x AY103656 (1-1665)

```

QY 9 SerGInGInValGlyGInMetYrAspLeuValThrProLeuLeuAsnSerValAlaGly 28
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 382 ACTGATATGTTAAATAAATACTATGATCTTGCCACTGACTTCTATGATGATGTTGGGT 441
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 29 GlyProCysAlaIleHisGlyYrTrpGluAsnAspGlyArgAlaSerTrpGlnGln 48
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 442 GAATCCTTCACCTTGGCTCAC---AGATGGAATGGAGAA-----TCCTTAGCTGAA 489
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 49 AlaAlaAspArgLeuThrAspLeuValAlaGluArgThrValIleuAspGlyValAlaArg 68
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 490 AGCATCAACGACATGAGCATTTTCTTGCCCTGCAGCTTGCTTGAAACCGAATGAAAG 549
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 69 LeuLeuAspValGlyCysGlyThrGlyGlnProAlaLeuArgValAlaArgAspAsnAla 88
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 550 GTTTTGAATGTGGGCTGCGAATGAGTGAGCACCTGAGAAATTCGAAAGATTAGCTCA 609
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 89 IlegInIleThrGlyIleThrValSerGlnValGlnValAlaIleAlaAlaAspCysAla 108
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 610 ACTTCAGTTACCGGATTTGAATTAACACGAAATTAACGAGGGAAGAGACTCAAC 669
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 109 ArgGluArgGlyLeuSerHisArgValAspPheSerCysValAspAlaMetSerLeuPro 128
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 670 CGTTTAGCGAAGAAATTAGGAAACATGTATTTTCTCAAGCGGACCTTCATTAAGATGCCG 729
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 129 TyrProAspAsnAlaPheAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 148
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 730 TTCATGCAACACTTTTGTATGCTGTTTACCCATTTAGGCAACATGATGACACCTGAT 789
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 149 ProAspArgAlaIleArgGluIleLeuArgValIleuValLeuValProGlyGlyIleLeuGlyVal 168
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```





```

Db      552 TACAGACAGATTACACATCATTAACACCCGGCGGTGTTTGGTGTATACGANTGCGC 493
Qy      173 LysArgLysAla---GlyGlyMetProVal----- 182
Db      492 ATGACCGACGCTTATGATATACAAACACCCGCTCCAGATGATCAGACTGGTATCGAA 433
Qy      183 SerGlyAspArgTrpProThrGlyLeuAlaGlyCysLeuAlaGluInLeuLeuGluSer 202
Db      432 CAAGGAGATGGAATTTGCAACATGATGTCAGG-----GCAACCGAAGCCCTTGATGCC 382
Qy      203 LeuArgAlaAlaGlyPheGlyLeuAspTrpGluAspValSerSerArg 219
Db      381 TTAAAGCCGCGCTTTCAGACTCATCAAGCTGAGAGATCTACCAACCGT 331

RESULT 14
AJ431900      634 bp      mRNA      linear      EST 15-MAR-2002
LOCUS      AJ431900      S00007      Hordeum vulgare cDNA clone S0000700049603FL, mRNA
DEFINITION      sequence.
ACCESSION      AJ431900      GI:19520352
VERSION      AJ431900.1
KEYWORDS      EST.
SOURCE      Hordeum vulgare
ORGANISM      Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 634)
Saren,A.-M., Tanskanen,J., Paulin,L. and Schulman,A.H.
REFERENCE      Unpublished (2002)
AUTHORS      Contact: Schulman AH
JOURNAL      Institute of Biotechnology
COMMENT      University of Helsinki
P.O.Box 56 (Wilkinlaari 6A), University of Helsinki FIN-00014,
Finland.

FEATURES
     source          1..634
                     /organism="Hordeum vulgare"
                     /mol_type="mRNA"
                     /db_xref="taxon:4513"
                     /clone="S0000700049603FL"
                     /dev_stage="shoot"
                     /clone_lib="S00007"
                     /note="2,-3,-4-days after germination"

ORIGIN
Alignment Scores:
Pred. No.:      3,81e-15      Length:      634
Score:          236.00      Matches:      65
Percent Similarity: 45.37%      Conservative: 33
Best Local Similarity: 30.09%      Mismatches: 86
Query Match:    16.34%      Indels:      32
DB:              1          Gaps:      6

US-10-069-353a-8 (1-275) x AJ431900 (1-634)
Qy      16 TyrAspLeuValThrProLeuLeuAsnSerValAlaGlyGlyProCysAlaIleHis 35
Db      8 TATGATCTTGTCTACTACTCTCTACGATATGCGGTGATTCCTCCACTTGGCTCAC 67
Qy      36 GlyTrpTrpGluAsnAspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAsp 55
Db      68 ---AGGTGAGAGTGAGAA-----TCTTAAAGGAAAGTATCAACCGCAGACAC 115
Qy      56 LeuValAlaGluArgThrValLeuAspGlyGlyValArgLeuAspValGlyCysGly 75
Db      116 TTCTGTGCGCTTACGAGCTTGAAGTAAACGAGGATGAAGTTTGGATGTCGGCTGTGGA 175
Qy      76 ThrGlyGlnProAlaLeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThr 95
Db      176 ATAGCGCGCGCGCTTAAAGAAATTGCGAGATTAGCTCCACCTCAATTACTGATGAAC 235

```

```

Qy      96 ValSerGlnValGlnValAlaIleAlaAspCysAlaArgGluArgGlyLeuSerHis 115
Db      236 AACACGACTTCAATATATAGGGAAGAGAGCTTATCGCTGTAGACTTGTGCA 295
Qy      116 ArgValAspPheSerCysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAsp 135
Db      296 ACTTGATATTGTCAGAGCAGACTTCAATGAAGATGCAATCTCTGATTAACACTTTGAT 355
Qy      136 AlaAlaTrpAlaMetGlnSerLeuLeuGluMetSerGluProAspArgAlaIleArgGlu 155
Db      356 GGTGTGTTTCCATTGAGGACCAATGCCAGCGCTGATCCGCTGCTCTATTAAGAG 415
Qy      156 IleLeuArgValLeuLeuProGlyGlyIleLeuGlyValThrGlu----- 170
Db      416 ATCTACCGTGTATTAAACCCGGGACAGTGTTCGTGTATATGATGTGATGATACCGAT 475
Qy      171 -----ValValLysArg-----GluAlaGlyGly 178
Db      476 CACTATGATCCAAACATCAACCAACAGAGATTAAGATGAATGAGCTGTGATAC 535
Qy      179 GlyMetProValSerGlyAspArgTrpProThrGlyLeuArgIleCysLeuAlaGluIn 198
Db      536 GGCGTCCCA-----GATATCAAGAACTACTGGCGCA 565
Qy      199 LeuLeuGluSerLeuArgAlaAlaGlyPheGlyLeuLeuAspTrpGlu 214
Db      566 TGTCTTCAAGCTTAAAGATCTGGGTTTGAGGTTATT---TGGGAT 610

RESULT 15
BJ271067      650 bp      mRNA      linear      EST 09-APR-2002
LOCUS      BJ271067      Y. Oghihara unpublished cDNA library, Wh_oh
DEFINITION      aestivum cDNA clone whn23p14 5', mRNA sequence.
ACCESSION      BJ271067      GI:20096630
VERSION      BJ271067.1
KEYWORDS      EST.
SOURCE      Triticum aestivum (bread wheat)
ORGANISM      Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 650)
Oghihara,Y. and Murai,K.
REFERENCE      Expressed genes in Triticum aestivum
AUTHORS      Unpublished (2002)
JOURNAL      Contact: Tadasi Shin-i
COMMENT      Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
     source          1..650
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="whn23p14"
                     /tissue_type="pistil at heading date"
                     /dev_stage="Feekes' scale 10.5"
                     /clone_lib="Y. Oghihara unpublished cDNA library, Wh_oh"

ORIGIN
Alignment Scores:
Pred. No.:      6.51e-15      Length:      650
Score:          234.00      Matches:      66
Percent Similarity: 44.05%      Conservative: 34
Best Local Similarity: 29.07%      Mismatches: 95
Query Match:    16.20%      Indels:      32
DB:              4          Gaps:      6

US-10-069-353a-8 (1-275) x BJ271067 (1-650)

```

OY		5	GlyAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThrProLeuLeuAsn	24
Dd		1	GGAGAGAActTAACCCGATATGGTTAAATAACTAATGAATCTGTACTACTATTAG	60
OY		25	SerValAlaGlyGlyProCysAlaIleHisGlyTyrTrpGluAsnSpgIArgAla	44
Dd		61	TATGGCTGGGGTAGAGCTTCACCTTGCTCAC---AGGTGGAAATGGAGAAA-----	108
OY		45	SerTPrgInGlnAlaAlaAspArgLeuThrAspLeuValAlaGluLysrghThrValLeuAsp	64
Dd		109	TCCTTAAAGGAAAGCAATCAAAGCGACAGACATCTTGCTTAAAGCTTGAGTTGAAA	168
OY		65	GlyGlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAlaLeuArgValAla	84
Dd		169	CCAGGAGATGAAGGTTTGGATGTCGGCTGTGGAAATAGGGGGGCATTAAAGAAATTCG	228
OY		85	ArgAspAsnAlaIleGlnIleThrGlyIlethrValSerGlnValGlnValAlaIleAla	104
Dd		229	AGATTAGCTCCACCCTCAGCTTACTGATTGAACCAACAAGCATCCAGATCAATAGGGGA	288
OY		105	AlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSerCysValAspAla	124
Dd		289	AAGCGCGCTTAATCCGTCGGTAGACTTGGCGCACTTGTTGATTTGTCAAGCAGACTTC	348
OY		125	MetSerLeuProTyrProAspAsnAlaPheAspAlaAlaITrpAlaMetGlnSerLeu	144
Dd		349	ATGAAGATGCCATCTCTCATTAACATTTTGATGCTGTTATCCATCGAGCAACATGC	408
OY		145	GluMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeuLysProGlyGly	164
Dd		409	CAGCACTGATGTCGGTTGGCTGCTCAAGAGAGATCTACCGTGTATTAAACCCGGGACG	468
OY		165	IleLeuGlyValThrGlu-----ValVal	172
Dd		469	TGTTTTCGTATPATAGAGTGTGCATTACCGATCACTATGATGCCAAACAATGACACCC	528
OY		173	LysArg-----GluAlaGlyGlyGlyMetProValSerGlyAspArgTrp	187
Dd		529	AAGAGAGATTAAAGATGAATTAAGACTTGGGAATGATGCCA-----	570
OY		188	ProTrhGlyLeuArgIleCysLeuAlaGluGlnIleLeuGlnLeuSerLeuArgAlaIleAgly	207
Dd		571	-----GARATCAGAAGTACTCGGCATATGCTTCACAGCTGTAAAGATGCTGGG	618
OY		208	PheGluIleLeuAspTrpGlu	214
Dd		619	TTTTAGGTATAT---TGGGAC	636

Search completed: August 29, 2005, 22:18:49  
Job time : 3551 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 29, 2005, 19:39:44 ; Search time 228 Seconds  
(without alignments)  
1973.579 Million cell updates/sec

Title: US-10-069-353a-8

Perfect score: 1444

Sequence: 1 VLEGAPTSQGVQGMVDTLT.....YERYADMGVAILTARKPVG 275

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE:frame+p2n.model -DEV:x1h  
-O:/cgnt2\_1/uspro.epool/US10069353/runat\_26082005\_172054\_14693/app\_query.fasta\_1.455  
-DB:issued\_patents.na -OPMT:fastap -SUFFIX:mi -MINMATCH=0.1 -LOOFCU=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=dlcsm62 -TRANS=human40.cdt  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USR=US10069353 @CGN 1.1 69 @runat\_26082005\_172054\_14693 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*

1: /cgnt2\_6/ptodata/1/ina/5A.COMB.seq:\*\n2: /cgnt2\_6/ptodata/1/ina/5B.COMB.seq:\*\n3: /cgnt2\_6/ptodata/1/ina/6A.COMB.seq:\*\n4: /cgnt2\_6/ptodata/1/ina/6B.COMB.seq:\*\n5: /cgnt2\_6/ptodata/1/ina/PTCUS.COMB.seq:\*\n6: /cgnt2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1444	100.0	80161	3	US-09-036-987A-1
2	1444	100.0	80161	4	US-09-370-700-1
3	1444	100.0	80161	4	US-09-603-207-1
4	435.5	30.2	80161	3	US-09-036-987A-1
5	435.5	30.2	80161	3	US-09-370-700-1
6	435.5	30.2	80161	4	US-09-603-207-1
7	379	26.2	6085	3	US-09-029-603-4
8	376.5	26.1	852	4	US-09-266-965-36
9	376.5	26.1	53500	4	US-09-266-965-76
10	345	23.9	828	4	US-09-266-965-35
11	320.5	22.2	2122	3	US-09-029-603-1
12	241.5	16.7	1444	4	US-09-885-723-21

13	216	15.0	1164	4	US-09-248-796A-3873
14	216	15.0	1400	3	US-09-041-718-1
15	210.5	14.6	1990	4	US-09-118-637A-3
16	206	14.3	954	4	US-09-118-637A-1
17	200.5	13.9	1421	4	US-09-885-723-22
18	198	13.7	441529	3	US-09-103-840A-1
19	196	13.6	957	4	US-09-382-906A-1
20	195.5	13.5	1152	4	US-09-614-221A-523
21	193.5	13.4	4403765	3	US-09-103-840A-2
22	187.5	13.0	1369	4	US-09-439-554-29
23	181.5	12.6	1062	4	US-09-902-540-3096
24	181.5	12.6	15499	4	US-09-902-540-1140
25	179.5	12.4	759	4	US-09-583-110-637
26	179.5	12.4	780	4	US-09-107-433-1646
27	179.5	12.4	1249	4	US-09-885-723-20
28	177.5	12.3	3236	3	US-08-961-527-222
29	170.5	11.8	9377	3	US-09-221-017B-1002
30	168.5	11.7	1131	4	US-09-252-991A-2322
31	168.5	11.7	1215	4	US-09-252-991A-2513
32	168.5	11.7	1215	4	US-09-252-991A-2595
33	162.5	11.3	783	4	US-09-489-039A-5480
34	162	11.2	6599	4	US-09-902-540-724
35	161	11.1	597	4	US-09-252-991A-16384
36	161	11.1	1866	4	US-09-252-991A-16213
37	154	10.7	1830121	4	US-09-557-884-1
38	154	10.7	1830121	4	US-09-643-990A-1
39	153.5	10.6	696	4	US-09-902-540-3923
40	153.5	10.6	10263	4	US-09-902-540-1024
41	152.5	10.6	747	4	US-09-902-540-9423
42	152.5	10.6	11358	4	US-09-902-540-1075
43	152	10.5	20303	4	US-09-902-540-1183
44	151.5	10.5	4403765	3	US-09-103-840A-2
45	151.5	10.5	4411529	3	US-09-103-840A-1

#### ALIGNMENTS

RESULT 1  
US-09-036-987A-1  
Sequence 1, Appli Application US/09036987A  
Patent No. 6143526  
GENERAL INFORMATION:  
APPLICANT: Baltz, Richard H.  
APPLICANT: Broughton, Mary C.  
APPLICANT: Crawford, Kathryn P.  
APPLICANT: Maddux, Krishnamurthy  
APPLICANT: Merlo, Donald J.  
APPLICANT: Treadway, Patti J.  
APPLICANT: Turner, Jan R.  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
TITLE OF INVENTION: Production  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dow AgroSciences LLC Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/036,987A  
FILING DATE: 09-MAR-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stuart, Donald R  
REGISTRATION NUMBER: 28,479  
REFERENCE/DOCKET NUMBER: 50,608

```
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (317)337-4816
/ TELEFAX: (317)337-4847
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/   LENGTH: 80161 base pairs
/   TYPE: nucleic acid
/   STRANDEDNESS: double
/   TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

Alignment Scores:
Pred. No.: 4,54e-160      Length: 80161
Score: 1444.00           Matches: 275
Percent Similarity: 100.00%  Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 100.00%      Gaps: 0
DB: 3

US-10-069-353A-8 (1-275) x US-09-036-987A-1 (1-80161)

QY 1 ValLeuProGlyGlyAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20
Db 20168 GTGTTCACAGGTGGCGCACCAATCGCAGAGTTGGGCAATGATGACCTGTGCAGC 20227
QY 21 ProLeuLeuAsnSerValAlaGlyGlyProCysAlaIleHisGlyTyrTrpGluAsn 40
Db 20228 CCGTTGCTGAATCGGTCTCCGGCGGCGCCCTTCGCGCATCCACACCGCTACTCGGAGAAC 20287
QY 41 AspGlyArgAlaSerTrpGlnGlnAlaAspArgLeuThrAspLeuValAlaGluArg 60
Db 20288 GACGGGCGGCGCTTCCTGCGCAGCAGCGCCGACCGGCTCACCGACTTGTCCCGAACCG 20347
QY 61 ThrValLeuAspGlyGlyValArgLeuAspValGlyCysGlyThrGlyGlnProAla 80
Db 20348 ACCGAGCTCGATGGCGCGCTGACCTGATGATGGGGTGGCGGTACCGGACACCGAGG 20407
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
Db 20408 CTGGCGGTGGCGGCGGACCAACCGATCCGATCACCGGATCACCGTCCAGGAGTGCA 20467
QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
Db 20468 GTGGCGATCGCGCTGATGGCCGACGCGAACCGGACTTAAGCCACCTTGTGCCGAACTTCG 20527
QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
Db 20528 TCCGTCGATGCAATGCTCCCTGCGGACCGGACAAATGCTTTCGACGCGCGCTGGGCGCATG 20587
QY 141 GlnSerLeuGluGlnMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160
Db 20588 CAGTCGCTGTGGAGATGTCGGAACCGGACCGTGCATCCGGGAATCTTCTGAGTACTC 20647
QY 161 LysProGlyGlyIleLeuGlyValThrGluValValValArgGluAlaGlyGlyGlyMet 180
Db 20648 AAACCCGGGTGCATCTCGCGCTCACCGAGTCTGCAACGGAAGCGGCGCGCGGATG 20707
QY 181 ProValSerGlyAspArgTrpProThrGlyLeuArgIleCysLeuAlaGluGlnLeu 200
Db 20708 CCGGTGTCGGGGGACAGGTGGCGGACCGGCTTCGGATGCTGCTGAGGAACTTCTG 20767
QY 201 GlnSerLeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220
Db 20768 GAATCGCTGCGTGACGCGGAGTTTCGAGATCTTCGATTGGGAGAGACGTGTCGCGAGACC 20827
QY 221 ArgTyrPheMetProGlnPheAlaGluGluLeuAlaAlaHisGlnHisGlyIleAlaAsp 240
Db 20828 CGGTACTTCATGCCCAAGTTCGCCGAAGAGCTCGTGGCGCACACACGAGGATCCCGGAC 20887
QY 241 ArgTyrGlyProAlaValAlaGlyTrpAlaAlaAlaValCysAspTrpGluSerTyrAla 260
Db 20888 AGGTACGGCGCGGCTGTGCGCGGCTGGCGCGCGGCTGTGCGATTATGAGAAATATGCC 20947
```

```
QY 261 HisAspMetGlyTyrAlaIleLeuThrAlaArgLysProValGly 275
Db 20948 CACGACATGGGCTATGCAATTCGACGCGCGGAAACCGGCTCGGC 20992

RESULT 2
US-09-370-700-1
/ Sequence 1, Application US/09370700
/ Patent No. 6274350
/ GENERAL INFORMATION:
/ APPLICANT: Baltz, Richard H
/ APPLICANT: Broughton, Mary C
/ APPLICANT: Crawford, Kathryn P
/ APPLICANT: Madhuri, Krishnamurthy
/ APPLICANT: Treadway, Patil J
/ APPLICANT: Turner, Jan R
/ APPLICANT: Waldron, Clive
/ TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
/ FILE REFERENCE: 50489 DIV1
/ CURRENT APPLICATION NUMBER: US/09/370,700
/ CURRENT FILING DATE: 1999-08-09
/ EARLIER APPLICATION NUMBER: US 09/36987
/ NUMBER OF SEQ ID NOS: 39
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 1
/ LENGTH: 80161
/ TYPE: DNA
/ ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Alignment Scores:
Pred. No.: 4,54e-160      Length: 80161
Score: 1444.00           Matches: 275
Percent Similarity: 100.00%  Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 100.00%      Gaps: 0
DB: 3

US-10-069-353A-8 (1-275) x US-09-370-700-1 (1-80161)

QY 1 ValLeuProGlyGlyAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20
Db 20168 GTGTTCACAGGTGGCGCACCAATCGCAGAGTTGGGCAATGATGACCTGTGCAGC 20227
QY 21 ProLeuLeuAsnSerValAlaGlyGlyProCysAlaIleHisGlyTyrTrpGluAsn 40
Db 20228 CCGTTGCTGAATCGGTCTCCGGCGGCGCCCTTCGCGCATCCACACCGCTACTCGGAGAAC 20287
QY 41 AspGlyArgAlaSerTrpGlnGlnAlaAspArgLeuThrAspLeuValAlaGluArg 60
Db 20288 GACGGGCGGCGCTTCCTGGCAGCAGCGCGGACCGGCTCACCGACTTGTGCCGAACTTCG 20347
QY 61 ThrValLeuAspGlyGlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80
Db 20348 ACCGAGCTCGATGGCGGCGGTTGCACTGCTCATGTGGGGTCCGGTACCGGACCAACGCG 20407
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
Db 20408 CTGGCGGTGGCGGCGGCGGACCAACCGATCCAGATCCCGGATCCAGTACGCAAGTGTCA 20467
QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
Db 20468 GTGGCGATCGCGCTGATGGCGCAGCGGACCGGACTTAAGCCACCGGAGTGAATCTTCG 20527
QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
Db 20528 TCCGTCGATGCAATGCTCCCTGCGGACCGGACAAATGCTTTCGACGCGCGCTGGGCGCATG 20587
QY 141 GlnSerLeuGluGlnMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160
Db 20588 CAGTCGCTGTGGAGATGTCGGAACCGGACCGGCTCCATCCGGGAATCTTCTGAGTACTC 20647
```





```
/ FILING DATE: 09-MAR-1998
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Stuart, Donald R
/ REGISTRATION NUMBER: 28,479
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (317)337-4816
/ TELEFAX: (317)337-4847
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 80161 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

Alignment Scores:
Pred. No.: 1,86e-39 Length: 80161
Score: 435.50 Matches: 108
Percent Similarity: 51.67% Conservative: 47
Best Local Similarity: 36.00% Mismatches: 96
Query Match: 30.16% Indels: 49
DB: Gaps: 11

US-10-069-353A-8 (1-275) x US-09-036-987A-1 (1-80161)
QY 7 Prothrsrserglnlnvalglglnmetlrraspleuvalthrproleuleuanserval 26
DB 13562 CCGACCGCGGATCAGGTGAGTGCATCTTCGATGCGTTG-----GCG 13521
QY 27 Alaaglylprocyalailehshisgllytyrrp-----gluasnasp 41
DB 13520 CACGGGCGTCCC-----CTGCACACCGGTTACTGGGGCGGGGCGGATCGGAGATGCC 13467
QY 42 Glyargalasertypglnlnlnlaaaspargleuthraspleuvalaagluargthr 61
DB 13466 GGTGCACACCGTGTGCGATGCTGCGCACCAATGACCGACTGTTCATGCACAGGCC 13407
QY 62 Valleunaspglylvalargleuleuaspvalglycygslylthrglylnprolaieu 81
DB 13406 GCGCTCCGTCGCGAGCGGACCTGTTGCACCTGGGCTGCGGCAATGGGACGCCGTGATGC 13347
QY 82 Argvalalaargaspasnaalaileglnlthrglylththrvalserglnlnval 101
DB 13346 CCGTGGCGCATGGCGCCAGCGCGCTTCGACGACCGGAATACCGTGAAGCCCGACATCTC 13287
QY 102 Alailealaalaaspcysalaargluarglyleuserhisaargvalasphesercys 121
DB 13286 GCGCGCGCACACGAGCTGCCCAACGACCGGACTGGCGCGGACGTTCGATTCGATCTA 13227
QY 122 ValaaspalametserleuProtyrProaasnaalaphesalalaatrpalametgln 141
DB 13226 GTGCAGCGCGCCACGATGCTCCCAACCGGACGCTTCACGCGCGCATGGCGCATGACG 13167
QY 142 Serleuleuagluuserglnuproaspargalalailearglnlthrglylnleuargvalaieuys 161
DB 13166 TCCGTCGTCACATGCTGCACACGAGCGCGCGGATCCGCGAGGTCCACCGAATCTTGGA 13107
QY 162 Proglglyl-----lleglnlyl-----Valthrglnvalaallysarglnaagly 177
DB 13106 CCGCGCGCGCGGTTGCTCTCGAGACATCATCCTCGGGTTGCACTCCCGGAAAGATAC 13047
QY 178 GlyglymetProvalserglyaspargtrpprothrglyleuarglilecysleuaglu 197
DB 13046 GCGGCGGTT-----TGG-----ACGGGACGACCGCCCATCTTGAAC 13008
QY 198 GlnleuleuGluSerleuargalaaglylphaglnlthrglylnleuaspdrglnuaspvalser 217
DB 13007 AGCTTCACGCGCGCTGTCAGCGAAGCGGGTTGAGATTCCTGAAGTACCGACCTCAGC 12948
QY 218 SerargthrargtyrphenetpProglInphealaglulnleu----- 231
```

```
DB 12947 GCACAGACAGATGATGCTCTCTGATCTGCACGAGTGTCTCCGGAATCGATGAG 12888
QY 232 -----Alalaahisglnhs-----Glylle 238
DB 12887 CTCGCCGCGTCGACCTCGCGGCTGTGCGCACCTTCACGACACGCTACTTGGAGACATC 12828
QY 239 AlaasparglytyrProalaalaaglylthrpalaalaalavalcyaspytyrGlnuys 258
DB 12827 GCGGGGAAGCACGACCGCGGACACGACAGCTGATGCGCGGGTTGGGAATACCGGAA 12768
QY 259 -----Tyralahisasp-----Metglytyralaileleuthralaarglys 272
DB 12767 CATCCGATTAACGACGAAGAGGAAAGCATGGGTTTCATGCTCTCAGGCTCGAAG 12708

RESULT 5
US-09-370-700-1/c
/ Sequence 1, Application US/09370700
/ Patent No. 6274350
/ GENERAL INFORMATION:
/ APPLICANT: Baltz, Richard H
/ APPLICANT: Broughton, Mary C
/ APPLICANT: Crawford, Kathryn P
/ APPLICANT: Madhuri, Krishnamurthy
/ APPLICANT: Treadway, Patil J
/ APPLICANT: Turner, Jan R
/ APPLICANT: Waldron, Clive
/ TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
/ FILE REFERENCE: 50489 DIV1
/ CURRENT APPLICATION NUMBER: US/09/370,700
/ CURRENT FILING DATE: 1999-08-09
/ EARLIER APPLICATION NUMBER: US 09/36987
/ EARLIER FILING DATE: 1998-03-09
/ NUMBER OF SEQ ID NOS: 39
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 1
/ LENGTH: 80161
/ TYPE: DNA
/ ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Alignment Scores:
Pred. No.: 1,86e-39 Length: 80161
Score: 435.50 Matches: 108
Percent Similarity: 51.67% Conservative: 47
Best Local Similarity: 36.00% Mismatches: 96
Query Match: 30.16% Indels: 49
DB: Gaps: 11

US-10-069-353A-8 (1-275) x US-09-370-700-1 (1-80161)
QY 7 Prothrsrserglnlnvalglglnmetlrraspleuvalthrproleuleuanserval 26
DB 13562 CCGACCGCGGATCAGGTGAGTGCATCTTCGATGCGTTG-----GCG 13521
QY 27 Alaaglylprocyalailehshisgllytyrrp-----gluasnasp 41
DB 13520 CACGGGCGTCCC-----CTGCACACCGGTTACTGGGGCGGGGCGGATCGGAGATGCC 13467
QY 42 Glyargalasertypglnlnlnlaaaspargleuthraspleuvalaagluargthr 61
DB 13466 GGTGCACACCGTGTGCGATGCTGCGCACCAATGACCGACTGTTCATGCACAGGCC 13407
QY 62 Valleunaspglylvalargleuleuaspvalglycygslylthrglylnprolaieu 81
DB 13406 GCGCTCCGTCGCGAGCGGACCTGTTGCACCTGGGCTGCGGCAATGGGACCGCTGATC 13347
QY 82 Argvalalaargaspasnaalaileglnlthrglylththrvalserglnlnval 101
DB 13346 CCGTGGCGCATGGCGCGCGGCTTCGACGACCGGAATACCGTGAAGCCCGACATCTC 13287
QY 102 Alailealaalaaspcysalaargluarglyleuserhisaargvalasphesercys 121
```

```

Db      13286 GCCCGCCGACGAGCTCGCAGACGAGACCGGAGCTGGCGAGCTTGATGATCTGA 13227
Qy      122 ValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTTPAlaMetGln 141
Db      13226 GTCGACGCGGCCCGGACGCTGCTTACCCGAGCGTTTCTTTCAGGCCCGCCATGGGGAGTGACG 13167
Qy      142 SerLeuLeuGluMetSerGluProAspAspAlaAlaArgGluIleLeuArgValLeuLys 161
Db      13166 TCCGTGTGCAGATCGTGGACGACGCGCCGCGCATCCGACAGGTCCACCGAATCTCTGGAA 13107
Qy      162 ProGlyGly-----IleLeuGly-----ValThrGluValValLysArgGluAlaGly 177
Db      13106 CCCGCGCGCGCGGTTCCGTCTCCGAGACATCATCTCGGTTGCACTCCCGGAAAGATAC 13047
Qy      178 GlyGlyMetProValSerGlyAspArgTTPProThrGlyLeuArgIleCysLeuAlaGlu 197
Db      13046 GCGCGCGGTT-----TGG---ACGGGACGACCGCCCATCTTGAAC 13008
Qy      198 GlnLeuLeuGlnSerLeuArgAlaAlaGlyPheGlnIleLeuAspTTPGluAspValSer 217
Db      13007 AGCTTCACGCGCGCTGTCAGCGGACCGGCTTGAGATCTCGAAGTCACCGACCTCAG 12948
Qy      218 SerArgThrArgTyrPheMetProGlnPheAlaGluGluLeu----- 231
Db      12947 GCACAGACCGAGTGCATGTCTCTCTGTAGTCAGACAGTTGCTCCGAAACCTCGATGAG 12888
Qy      232 -----AlaAlaHisGlnHis-----GlyIle 238
Db      12887 CTCGCGCGCGCTGACGCTGCGGCTGTCGACCTACGACGACGCTTACTTGGAGACATC 12828
Qy      239 AlaAspArgTyrGlyProAlaValAlaGlyTTPAlaAlaValCysAspTyrGluLys 258
Db      12827 GCGGCGAAGCAGCAGCAGCGGAGCAGCAGCTGATCGCGGTTGGCGAATACCGGAA 12768
Qy      259 -----TyrAlaHisAsp-----MetGlyTyrAlaIleLeuThrAlaArgLys 272
Db      12767 CATCCGGATTACCGCAGAAACGAGAAAGCATGGGTTTCACTGCTCTCGACGCTCGAAG 12708

```

```

RESULT 6
US-09-603-207-1/c
; Sequence 1, Application US/09603207B
; Patent No. 6521406
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/603,207B
; EARLIER FILING DATE: 2000-06-23
; EARLIER APPLICATION NUMBER: 09/370,700
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
US-09-603-207-1

```

```

Alignment Scores:
Pred. No.: 1,86e-39 Length: 80161
Score: 435.50 Matches: 108
Percent Similarity: 51.67% Conservative: 47
Best Local Similarity: 36.00% Mismatches: 96
Query Match: 30.16% Indels: 49
DB: 4 Gaps: 11

```

US-10-069-353a-8 (1-275) x US-09-603-207-1 (1-80161)

```

Qy      7 ProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThrProLeuLeuAsnSerVal 26
Db      13562 CCGACCGCGGATCAGGTGAGATTCATCTTCATAGTGTTG-----GCG 13521
Qy      27 AlaGlyGlyProCysAlaAlaIleHisGlyTyrTP-----GluAsnAsp 41
Db      13520 CACGGGCGGTCC-----CTGCACACGGGTACTGGGCGGGGATGCGAGGATGCGC 13467
Qy      42 GlyArgAlaSerTTPGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArgThr 61
Db      13466 GGTGCCACACCGTGGGTGGATCTGCGACCAAGCTGACCTGTTCATCCACAGGCC 13407
Qy      62 ValLeuAspGlyGlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAlaLeu 81
Db      13406 GCGCTCCGTCCCGAGCGGACGCACTGTTCACACTCGGCTCGGCAAGGGGACGCCGTAGTC 13347
Qy      82 ArgValAlaAspAspAlaIleGlnIleThrGlyIleThrValSerGlnValGlnVal 101
Db      13346 CGTGGCGCATGGCCAGCGCGGCTTGAGTCAACCGAATCACCGTGAACGCCAGCATCTC 13287
Qy      102 AlaIleAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSerCys 121
Db      13286 GCGCGCCGACGAGGCTCGCCAAACGAGACCGGACTGGCCGCGAGCTTGAGTTGATCTTA 13227
Qy      122 ValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTTPAlaMetGln 141
Db      13226 GTCGACGCGGCCCGAGTGCCTTACCCGAGCGTTTCTTTCAGGCCCGCATGGCGATGACG 13167
Qy      142 SerLeuLeuGlnMetSerGluProAspAspAlaAlaArgGluIleLeuArgValLeuLys 161
Db      13166 TCCGTGTGCAGATCGTGGACGACGCGCCGCGCATCCGACAGGTCCACCGAATCTCTGGAA 13107
Qy      162 ProGlyGly-----IleLeuGly-----ValThrGluValValLysArgGluAlaGly 177
Db      13106 CCCGCGCGCGGTTCCGTCTCCGAGACATCATCTCGGTTGCACTCCCGAAGATGAC 13047
Qy      178 GlyGlyMetProValSerGlyAspArgTTPProThrGlyLeuArgIleCysLeuAlaGlu 197
Db      13046 GCGCGCGGTT-----TGG---ACGGGACGACCGCCCATCTTGAAC 13008
Qy      198 GlnLeuLeuGlnSerLeuArgAlaAlaGlyPheGlnIleLeuAspTTPGluAspValSer 217
Db      13007 AGCTTCACGCGCGCTGTCAGCGGACCGGCTTGAGATCTCGAAGTCACCGACCTCAG 12948
Qy      218 SerArgThrArgTyrPheMetProGlnPheAlaGluGluLeu----- 231
Db      12947 GCACAGACCGAGTGCATGTCTCTCTGTAGTCAGACAGTTGCTCCGAAACCTCGATGAG 12888
Qy      232 -----AlaAlaHisGlnHis-----GlyIle 238
Db      12887 CTCGCGCGCGCTGACGCTGCGGCTGTCGACCTACGACGACGCTTACTTGGAGACATC 12828
Qy      239 AlaAspArgTyrGlyProAlaValAlaGlyTTPAlaAlaValCysAspTyrGluLys 258
Db      12827 GCGGCGAAGCAGCAGCAGCGGAGCAGCAGCTGATCGCGGTTGGCGAATACCGGAA 12768
Qy      259 -----TyrAlaHisAsp-----MetGlyTyrAlaIleLeuThrAlaArgLys 272
Db      12767 CATCCGGATTACCGCAGAAACGAGAAAGCATGGGTTTCACTGCTCTCGACGCTCGAAG 12708

```

```

RESULT 7
US-09-603-4
; Sequence 4, Application US/09029603
; Patent No. 6210935
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Engel, Natalie
; APPLICANT: Bietenhader, Jurg
; APPLICANT: Toupet, Christine
; APPLICANT: Possiech, Andreas
; TITLE OF INVENTION: Staurosporin Biosynthesis Gene Clusters
; FILE REFERENCE: 4-20555/A/PCT

```



```

Db      139 GACAGCCAGGTGCGCTGGCCGACGACCGACCGGCTCAGCATGATGCGGACGG 198
Qy      61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCysGlyThrGlyInProAla 80
Db      199 CTGGCATGCGGCGCGCTCCGCGCTGCTGAGACCTGCGGCTGCGGCTGCGGACCGCGGCG 258
Qy      81 LeuArgValAlaArgAspAlaAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
Db      259 GATCGCATGCGCGGCTGCGGCGGACGATGTCAGGCGCATGTCCTCGGTGAGCCATGAGCGAG 318
Qy      101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
Db      319 GTCTGCGGCGGCGACCGCGCTGCGGCGGACGAGCGCGGCTGCGGCGGCGGCTTCCAG 378
Qy      121 CysValAspAlaMetSerLeuProTyrProAspAspAlaPheAspAlaAlaIleProAlaMet 140
Db      379 CGGCGCGACGCGATGACCTCCCTTCGAGAGCAGACGATTCGACCGCGCTCATCGCGCTTC 438
Qy      141 GlnSerLeuLeuGlnMetSerGluProAspArgAla-----IleArgGlnIleLeuArg 158
Db      439 GAATCGATCATCCACATG-----CCGACCGCGCGCCAGGTGCTGCGCCCGCGCGCG 492
Qy      159 ValLeuLeuProGlyGlyIleLeuGlyValThrGlyValValValysArg----- 174
Db      493 GTCTGCGGCGCGCGGCGCGCTGCTGCTGCTCAGCCGACTTCTTGAGCGCGCGCGCTCGCG 552
Qy      175 ---GluAlaGlyValGlyMetPro-----ValSer 183
Db      553 CCGGAGGCGGCGCGCGCGCTGCGGCGCTACCTCCAGACTTCATGATGACATGATGACG 612
Qy      184 GlyAspArgTrpProThrGlyLeuArg-----IleCys 194
Db      613 GCCGAGCGGCTGCTCCCTGCTGCTGCGGCGGCGCGGCTGCTGCTGAGAGATTCTTCGAC 672
Qy      195 LeuAlaGlnGlnLeuLeuGlnSerLeuArgAlaAlaGlyPheGlnIleLeuAspTrpGlu 214
Db      673 ATAGGACCAAGACCTTGGAGAAAGACC-----TTCAAGCTGCTCTGCG---GAG 717
Qy      215 AspValSerSerArgTrpArgTyrPheMetProGlnPheAlaGlnGlnLeuAlaAlaHis 234
Db      718 CGCATCAACTCTCGAAGACGAGGCTGAGACGACGATTCGCGGAGGAGATG---GTGAC 774
Qy      235 GlnHisGlyIleAlaAspArgTyrGlyProAlaValAlaGlyTrpAlaAlaValCys 254
Db      775 CAGTTGCAACCCCGCGGACCTCGTCGCGC----- 801
Qy      255 AspTyrGlnValSerTyrAlaHisAspMetGlyTyrAlaIleLeuThrAlaArgLysPro 273
Db      802 -----GTCAAGAGAGTTGCGCTATCTGCTGCTGCTGCGCGCGCGCG 843

```

# RESULT 9

```

US-09-266-965-76/c
; Sequence 76, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/09/266,965
; EARLIER FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 08/624,447
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: PCT/US94/11279
; EARLIER FILING DATE: 1994-10-06
; EARLIER APPLICATION NUMBER: US 08/133,963
; EARLIER FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: PasteSeq for Windows Version 3.0
; SEQ ID NO 76

```

```

; LENGTH: 53500
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
; US-09-266-965-76

Alignment Scores:
Pred. No.: 1,166-32
Score: 376.50
Percent Similarity: 50.84%
Best Local Similarity: 33.78%
Query Match: 26.07%
DB: 4 Gaps: 11

US-10-069-353a-8 (1-275) x US-09-266-965-76 (1-53500)

Qy      2 LeuProGlyValAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThrPro 21
Db      7560 CTCCCATGCGCTCCCTCCGCGAG-----AACCTGCACTTGCGCTACGGAGCTCCCG 7501
Qy      22 LeuLeuAspSerValAlaGlyValProCysAlaIleHisGlyTyrTrpGluAsn--- 40
Db      7500 CTGGAGCGCGCTCCCTCCGCGAG-----AACCTGCACTTGCGCTACGGAGCTCCCG 7447
Qy      41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
Db      7446 GACAGCGAGTCCCGCTGCGCGAGCGACCGACCGGCTCAGCATGATGCGCGAGCGG 7387
Qy      61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCysGlyThrGlyInProAla 80
Db      7386 CTGGCATGCGGCGCGCTCCGCGCTGCTGCTGAGAGCTTCCGCTGCGGCGCGCGCGG 7327
Qy      81 LeuArgValAlaArgAspAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
Db      7326 GTACGATGCGCGGCTCAGCGAGCGATGTCAGGAGCATTCGGTGAGCATGAGCAG 7267
Qy      101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
Db      7266 GTCTGCGGCGGCGAGCGCGCTGCGGCGGCGCGGCTGCGGCGGCGGCTTCCAG 7207
Qy      121 CysValAspAlaMetSerLeuProTyrProAspAspAlaPheAspAlaAlaIleProAlaMet 140
Db      7206 CGGCGCGAGCATGAGCTCCCTTCGAGAGCAGAGACTTCGAGCGCGCTCATCGCGCTTC 7147
Qy      141 GlnSerLeuLeuGlnMetSerGluProAspArgAla-----IleArgGlnIleLeuArg 158
Db      7146 GAATCGATCATCCACATG-----CCGACCGCGCGCCAGGTGCTGCGCGCGCG 7093
Qy      159 ValLeuLeuProGlyGlyIleLeuGlyValThrGlyValValValysArg----- 174
Db      7092 GTCTGCGGCGCGGCGCGCTGCTGCTGCTCAGCCGACTTCTTGAGCGGCGCGCGCTCGCG 7033
Qy      175 ---GluAlaGlyValGlyMetPro-----ValSer 183
Db      7032 CCGGAGGCGGCGCGCGCTCCAGCGCTACCTCCAGACTTCATGATGACATGATGACG 6973
Qy      184 GlyAspArgTrpProThrGlyLeuArg-----IleCys 194
Db      6972 GCCGAGCGGTACCTCCCTGCTGCGGCGGCGGCGGCTGCTGCTGAGAGATTCTTCGAC 6913
Qy      195 LeuAlaGlnGlnLeuLeuGlnSerLeuArgAlaAlaGlyPheGlnIleLeuAspTrpGlu 214
Db      6912 ATAGGACCAAGACCTTGGAGAAAGACC-----TTCAAGCTGCTCTGCG---GAG 6868
Qy      215 AspValSerSerArgTrpArgTyrPheMetProGlnPheAlaGlnGlnLeuAlaAlaHis 234
Db      6867 CGCATCAACTCTCGAAGACGAGGCTGAGACGACGATTCGCGGAGGAGATG---GTGAC 6811
Qy      235 GlnHisGlyIleAlaAspArgTyrGlyProAlaValAlaGlyTrpAlaAlaValCys 254
Db      6810 CAGTTGCAACCCCGCGGACCTCGTCGCGC----- 6784
Qy      255 AspTyrGlnValSerTyrAlaHisAspMetGlyTyrAlaIleLeuThrAlaArgLysPro 273

```



```

QY      45 SerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArgThrValLeuSer 64
      985 ACCATCGAGAGCGGAGTGAACCGGCTCAGGATGTTTATGATGAAACGGCTGAACGGGTAC 1044
QY      65 GtGtGtValArgLeuLeuAspValGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 84
      1045 GCCACCTCCACGCTCTCCACGCTCCGCGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1104
QY      85 ArgAspAsnAlaIleGlnIleThrGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 104
      1105 GCGCGACACGGGGGACGGGTGACCGGCGATCAGCATGACGAGGAGGAGGAGTACGAGCGCC 1164
QY      105 AlaAspCysAlaArgGluArgGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 124
      1165 AACCGGCTGGCGCGCGGAGCGGGGTCGCGGACCGGTCGCGTGTTCAGCATGCGGAGCGG 1224
QY      125 MetSerLeuProTyrProAspAsnAlaAspAspAlaAlaTyrAlaMetGlnSerLeuLeu 144
      1225 ATGAAACTGCGCTTCCCGGACGCGCTGTTGAGCGCGGTGATGCGGTGATGATCTGC 1284
QY      145 GtMetSerGluProAspArgAlaIleArgGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 164
      1285 CACATGCGCGACCGGACGAGGTGTTCACCGAGGTGTGCGGCGGTGCGCGCGCGCGCGG 1344
QY      165 IleLeuGtValThrGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 184
      1345 CGGATGCTCTCACCAGCATCTTCGAGCGCCAC-----CCGCGCAAGGCG 1389
QY      185 AspArgTyrProThrGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 196
      1390 GTTCGACACCGCGCGCATCGCAAGTTCGCGGACGTCGATGTCGACACCGCGGACATC 1449
QY      197 GtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 216
      1450 GACGACTACGTCGCGCTGTCGACCGCTCGCGGTGCGCGGTGCGCGGATGTCGTCGAC 1509
QY      217 SerSerArgThrArgTyrPheMetProGlnPheAlaGlnGluLeu-----Ala 232
      1510 ACCGAGACGACCAACGCTG-----CGCTCGCGGACGAGATCGGACGCTCGCGGCG 1560
QY      233 AlaHisGlnIleGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 242
      1561 GTCGAGAGCGCGCGCGGTCGATGACGAGGCACTTCGCTCGCGGACGACCTCTTC 1620
QY      243 GtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 261
      1621 AAGCCGTCGACCTCGCGGCG-----GTGAC 1647
QY      262 AspMetGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 273
      1648 GACTTCGCGTGCCTCTGTCGTCGCGGACGCGCGCG 1683
      1649 GACTTCGCGTGCCTCTGTCGTCGCGGACGCGCGCG

```

RESULT 12  
US-09-885-723-21  
Sequence 21, Application US/09885723

```

; Patent No. 6822142
; GENERAL INFORMATION:
; APPLICANT: Monsanto Company
; TITLE OF INVENTION: TRANSGENIC PLANTS CONTAINING ALTERED LEVELS OF STEROID COMPOUNDS
; FILE REFERENCE: WPC6783.1
; CURRENT APPLICATION NUMBER: US/09/885.723
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 1444
; TYPE: DNA
; ORGANISM: Nicotiana tabacum
; US-09-885-723-21

```

Alignment Scores:  
Pred. NO.: 7.29e-19  
Score: 241.50

Length: 1444  
Matches: 71

Percent Similarity: 49.15%  
Best Local Similarity: 30.08%  
Query Match: 16.72%  
DB: 4  
Gaps: 4

US-10-069-353a-8 (1-275) x US-09-885-723-21 (1-1444)

```

QY      9 SerGlnGlnValGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 28
      253 ACTGACATGTTAAACAAATACATATCTTTGACATAG-CTTCTACG-AAATCGCGTGGG 310
QY      29 GtProCysAlaIleHisGlnIleGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 48
      311 GAGAGTCATTCATTTTGACACC-AGTGGAAAGAGAA-----TCACCTCCAGAG 360
QY      49 AlaAlaAspArgLeuThrAspLeuValAlaGluArgThrValLeuAspGtGtGtGtGtGt 68
      361 AGCATTTAAAGGACATGACGACTTCTTCTGCTTGCACCTGGGATTTGAAACGAGCAAAAG 420
QY      69 LeuLeuAspValGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 88
      421 GTCTTGACGTGAGTGTGATGATTTGTTGGCGCGCTTAAGAGAAATGCTGATTCAGCTCT 480
QY      89 IleGlnIleThrGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 108
      481 ACATGATTAACAGCGCTCAACAAATGATATGATATCTTATGAGGACAGGTGTTGAC 540
QY      109 ArgGluArgGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 128
      541 CCGAAAGTGGATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY      129 TyrProAspAsnAlaAspAspAlaAlaTyrAlaMetGlnSerLeuLeuGtGtGtGtGtGt 148
      601 TTCCTGACAAATAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY      149 ProAspArgAlaIleArgGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 168
      661 CCATTGGATGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 720
QY      169 ThrGlu-----Val 171
      721 TATGAGTGTGTCATGACCGATCTTCAACCCCAATTAAGAAAGACAAACGATCAAG 780
QY      172 ValLeuArgGluAlaGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 191
      781 GCCGAATTTGAGTCTGGAATGCTCTCT----- 810
QY      192 ArgIleCysLeuAlaGlnGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 211
      811 GAGGTTAGATTGACACACAGTCCCTCGAAGCGCAACAGCTGTTTGAAGTTGTA 870
QY      212 AspTyrGluAspValSer-----SerArgThrArgTyrPheMetPro 225
      871 TGGACAAGATCTGCTGATGATCACTGTTCCATGATGATGCTT 918

```

RESULT 13  
US-09-248-796A-3873  
Sequence 3873, Application US/09248796A

```

; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248.796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074.725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096.409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3873
; LENGTH: 1164

```

	TYPE: DNA		
:	ORGANISM: Candida albicans		
US-09-248-796A-3873			
Alignment Scores:			
Pred. No.: 5,92e-16	Length: 1164		
Score: 216.00	Matches: 67		
Percent Similarity: 46.98%	Conservative: 42		
Best Local Similarity: 28.88%	Mismatches: 99		
Query Match: 14.96%	Indels: 24		
DB: 4	Gaps: 7		
US-10-069-353A-8 (1-275) X US-09-248-796A-3873 (1-1164)			
Oy	9 SerGInGlnValGlGlyInMetYrAspLeuValTrProLeuLeuAsnSerValAlAGly	28	
Db	241 TCCCAATTTGACTCATCATTATTAATTTAACTTAAAGGTGAA-----TATGGT	234	
Oy	29 GLYProCYsaIalIehShISgLy---TYrTPGIuaenAspGLyArGaLSeTrpGln	47	
Db	295 TGGGGTTTCATTCACATTTTTCACAGATAATTTAAAGGTGAA-----GCTTTTAGA	345	
Oy	48 GlnAlaIalAspArgLeuThrAspLeuValAlaGluArgThrValLeuAspGLyGlyVal	67	
Db	346 CAAGCTACTGCTAGACATGACAACATTCTTGCGCCATTAAAGTAATCTTAATGAAGAACATG	405	
Oy	68 ArgLeuLeuAspValGlyCySGlyThrglyGlnProAlaLeuArgValAlaArgAspAsn	87	
Db	406 AAAGTTTTAATGTTGGTTGGTGTAGAGTGGTCCGTGAGAGAAATCACAAAGATTACT	465	
Oy	88 AlaIleGlnIleThrglyIleThrValSerGlnValGlnValAlaIleAlaIAspCys	107	
Db	466 GATTGTGAATTTTGATTAATTAATTAATGAATATCAAATTTGAAAAGCTAAATCATTAT	525	
Oy	108 AlaArgGluArgglyLeuSerHisArgValAspPheSerCysValAspAlaMetSerLeu	127	
Db	526 GCATAAAAAAACCATTTAGATGCATAAATATCTTATGTTAAAGGTGATTTATGCAATG	585	
Oy	128 ProTyProAspAsnAlaPheAspAlaAlaTryPalamEtGlnSerLeuGlnMetSer	147	
Db	586 GATTTTAGAACCAAGAAATCATTTCCATGCTGTTTATGCCATTGAAGTACCGTTGATCTCCA	645	
Oy	148 GluProAspArgAlaIleArgGluIleLeuArgValLeuAspProGlyGlyIleLeuGly	167	
Db	646 GTTTTAGAAGAGACTTTATTACAGAAATTTTAAAGTTTGAAAACAGAGGTGATTTTGGT	705	
Oy	168 ValThrGluValVal-----LysArgGluAla	176	
Db	706 GTTTATGAATAGGGTCACTAGCATGAATAATACATGAAGAACTATGAAGAACTGTAAATT	765	
Oy	177 GlyGlyGlyMetProValSerGlyAspArgTrpProThrGlyLeuArgIleCysLeuAla	196	
Db	766 GCTTATAGGTATTGAAGATC--GATGATGGTATTCACAAAATAGTATTCTGTAAGATTGCT	822	
Oy	197 GluGlnLeuLeuGlnSerLeuArgAlaAlaGlyPheGluIle-----LeuAspTrp	213	
Db	823 GAACCAAGCTTTGAAAATATGT-----GGATTGAAAATTGAATATCAAAAAGATTTG	873	
Oy	214 GluAspValSerSerArgThrArgTrpPheMetPro	225	
Db	874 GCTGATGTTGATGATGATAAATTCCTTGATTTATTTCCA	909	
RESULT 14			
US-09-041-718-1	: Sequence 1/, Application US/09041718A		
	: Patent No. 6225075		
	: GENERAL INFORMATION:		
	: APPLICANT: Bard, Martin		
	: TITLE OF INVENTION: DNA encoding steroid methyltransferase		
	: FILE REFERENCE: 740.003US1		
	: CURRENT APPLICATION NUMBER: US/09/041,718A		
	: CURRENT FILING DATE: 1998-03-13		
	: NUMBER OF SEQ ID NOS: 9		

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-041-718-1

Alignment Scores:
Pred. No.:      7,81e-16      Length:      1400
Score:          216.00        Matches:     67
Percent Similarity: 46.98%    Conservative: 42
Best Local Similarity: 28.88% Mismatches:   99
Query Match:      14.96%      Indels:       24
DB:                3         Gaps:           7

US-10-069-353A-8 (1-275) x US-09-041-718-1 (1-1400)

OY      9      SerGlnGlnValGlyGlnMetTyrAspLeuValThrProLeuLeuAsnSerValAlaGly 28
      ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      346    TCCCAATTCATCATCATATTATTAATTTAGACCTTGACTTTTAGAA-----TATGT 399
      ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

OY      29      GlyProCysAlaIleHisHisGly--TYrrpGluAsnEspGlyArgAlaSerTrpGln 47
      ::::| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      400    TGGGGTTCTTCATTCATCATTTTCCAGATATATATAAGGTGA-----GCCTTAGA 456
      ::::| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

OY      48      GluAlaAlaAspArgLeuThrAspLeuValAlaGluArgThrValLeuAspGlyGlyVal 67
      ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      451    CAAGCTACTGCTGACATGACATGACATTTCTGGCCCCATAAAGAATCTTAATGAAAACATG 510
      ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

OY      68      ArgLeuLeuAspValGlyCyseGlyThrGlyGlnProAlaLeuAspValAlaArgAspAsn 87
      ::::::| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      511    AAAGTTTTAGATGTTGGTGTGTGTGTAGGTGCTCGTGTAGAGAAATCACAAAGTTTACT 570
      ::::::| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

OY      88      AlaIleGlnIleThrGlyIleThrValSerGlnValAlaAlaIleAlaAlaAspCys 107
      ::||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      571    GAATTGGAATTTGTTGGATTAAATAATAATGATTATCAATTTGAAAGAGCTAATCATAT 630
      ::||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

OY      108     AlArGgLuArGgLyLeuSerHIsArGyValAspPheSerCySvAlAspAlaMetSerLeu 127
      ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      631    GCTAAAAAATVACCATTAGATCATAAATTAATCTTAATGAAGGTATTTATGCAATG 690
      ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

OY      128     ProTyProAspAsnAlaPheAspAlaAlaTrpAlaMetGlnSerLeuLeuGluMetSer 147
      ::::| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      691    GAATTTGAACCGAATCATTCATGCTGTTATATGCAATGGAAGCTACCGTTACGCTCCA 750
      ::::| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

OY      148     GluProAspArgAlaIleArgGluIleLeuArgValLeuLysProGlyGlyIleLeuGly 167
      ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      751    GTTTTAGAAGGAGGTTTATTCACAATTTATAAGTTTGAACCAAGGTGGATTTTCGCT 810
      ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

OY      168     ValThrGluValVal-----LysArgGluAla 176
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      811    GTTTATGAATGGGTGATGATGATATAATACGATGAACATAATGAAGAACATCGTAAATT 870
      ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

OY      177     GlyGlyGlyMetProValSerGlyAspArgTrpProThrGlyLeuArgGlyIleCySLeuAla 196
      ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      871    GCTTATGATGATGAAGTCC--GGTATGCTATTCCAAATAATGATATTCTCGTAAAGTTGCT 927
      ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

OY      197     GluGlnLeuLeuGluSerLeuArgAlaAlaGlyPheGluIle-----LeuAspTrp 213
      ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      928    GAACAAGCTTTGAAAAAAGTT-----GGATTTGAAATTTGAATATCAAAAAGATTTG 978
      ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

OY      214     GluAspValSerSerArgThrArgTrpPheMetPro 225
      ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      979    GCTGATGTTGATGATGAATAATTCCTGGTATATATCCA 1014
      ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

RESULT 15
US-09-118-637A-3
; Sequence 3, Application US/09118637A
; Patent No. 6642434
; GENERAL INFORMATION:
; APPLICANT: Dellapenna, Dean
; APPLICANT: Shintani, David K.
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH TOCOPHEROL
```



```

/ TITLE OF INVENTION: METHYLTRANSFERASE
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Quarles & Brady
/ STREET: 1 South Plinkney Street
/ CITY: Madison
/ STATE: WI
/ COUNTRY: US
/ ZIP: 53701-2113
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/118,637A
/ FILING DATE:
/ CLASSIFICATION: 800
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Seay, Nicholas J.
/ REGISTRATION NUMBER: 27386
/ REFERENCE/DOCKET NUMBER: 920905, 90024
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 608-251-5000
/ TELEFAX: 608-251-9166
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1790 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 207..1253
/ US-09-118-637A-3

Alignment Scores:
Pred. No.: 5,13e-15 Length: 1790
Score: 210.50 Matches: 73
Percent Similarity: 39.46% Conservative: 45
Best Local Similarity: 24.41% Mismatches: 130
Query Match: 14.58% Indels: 52
DB: 4 Gaps: 10

US-10-069-353a-8 (1-275) x US-09-118-637A-3 (1-1790)
QY 1 ValLeuProGlyGlyAlaProThrSerGlnGln-----ValGlyGlnMet 15
DB 348 GTGCGTGTGGCGGCTGCTGCTACATCCATGAGCGCTAAGAAAGAAATAGCGGAGTTC 407
QY 16 TyrAspLeuValThrProLeuLeuAsnSerValAlaGlyProCysAlaIleHis 35
DB 408 TACATATAAATCTCGGTTTGTGGAGAGATTTGGGAGAT-----CATATGCATCAT 461
QY 36 GlyTyrTrpGluAsnAspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAsp 55
DB 462 GGCCTTTATGACCCCTGATTTCTTCTGTTCAACTTCTGATTCGTGCACAAAGAAAGCTCAG 521
QY 56 Leu---ValAlaGluArgThrValLeuAspGlyGlyVal----- 67
DB 522 ATCCGTATGATGTGAAGATCTCTCGTTTGGCCGGGTGTACTGATGTAAGAGAGAGAGAAA 581
QY 68 -----ArgLeuLeuAspValGlyCysGlyThrGlyGlnProAlaLeuArgValAla 84
DB 582 AAGATAAAGAAAGTAGAGATGTTGGGTGTGGATTTGAGGAAAGCTCAAGATATCTTCC 641
QY 85 ArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGlnValAlaIleAla 104
DB 642 TCTAATTTGGAGCTGATGATGATGATTAATCTCAAGCCCTGTTCAAGGCCAAGAGAGCC 701
QY 105 AlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSerCysValAspAla 124

```

```

DB 702 AATGATCTGGCGGCTGCTCAATCACTCTCTCATAGAGCTTCCTTCCAAATTGGGATGCG 761
QY 125 MetSerLeuProTyrTrpAspAsnAlaPheAspAlaAlaIleThrAlaMetGlnSerLeuLeu 144
DB 762 TTGGATTCAGCCATTCGAAGATGGAATAATTCAGTGTGTGTCATGAGAGTGCTGAG 821
QY 145 GluMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeuAspProGlyGly 164
DB 822 CATATGCTGACAAAGCCCAAGTTGTAAAGAGTTGTACGTGTCGGGCGCTCCAGAGGT 881
QY 165 IleLeuGlyValThrGluValValysArgGluAlaGlyGlyMetProValSerGly 184
DB 882 AGGATTAATATAGTGCATGTGCCATTAATCTATCTGCGGAGAGAACTTTG--- 938
QY 185 AspArgTrpProThrGlyLeu-----ArgIleCysLeuAla----- 196
DB 939 CAGCCGTGGAGCAAAACATCTTGACAAATCTGTAAAGACGTTCTATCTCCCGGCTGG 998
QY 197 -----GluGlnLeuGluSerLeuArgAlaAlaGlyPheGluIle 210
DB 999 TGCTCCACCGATGATTATGTCAATCTTCAATCCATTCCTCCAGGATTTTAAGTGT 1058
QY 211 LeuAspTrpGluAspValSerSerArgThrArgTyrPheMetProGlnPheAlaGluGlu 230
DB 1059 GCGGATGTGTCAGAG-----AACGTAGCTCTCTT- 1087
QY 231 LeuAlaAlaHisGlnHisGlyIleAlaAspArgTyrGlyProAlaValAlaGly----- 248
DB 1088 CTGGCCTGCGGTTATACGACATGATTAATGACGAAGGCGCTGTGTCTCTCTTCGTAG 1147
QY 249 TrpAlaAlaAlaValCysAspTyrGlySerGlyAlaHisAspMetGlyTyrAlaIle 267
DB 1148 TGG-----TATGAAAGTATTAAAGAGCATTCACATGCCATT 1186

```

Search completed: August 29, 2005, 22:25:30  
Job time : 406 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 29, 2005, 19:51:25 ; Search time 699 Seconds

(without alignments)  
2574.177 Million cell updates/sec

Title: US-10-069-353a-8

Sequence: 1 VLEPGAPTSQGVQGMVTLVT.....YEXYAHDMGYALITLTPKPVG 275

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 7331713 seqs, 3271544945 residues

Total number of hits satisfying chosen parameters: 14663426

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Command line parameters:

-MODEL=frame+ p2n.model -DEV=x1h  
-O/cgn2\_1/USPTO.apool/US10069353.r/unat.26082005.172055.14723/app.query.fasta\_1.455  
-DB=Published Applications NA -QFMT=fastap -SUPPLX=rnpb -MINMATCH=0.1  
-LOOPCU=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62  
-TRANS=human40.ccl -LIST=45 -DOCALLIGN=200 -THR SCORE=PCT -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10069353 @CGN 1.1 480 @runat.26082005.172055.14723  
-NCPU=6 -ICPU=3 -NO MAP -LARGESUBSTR -NEG SCORES=0 -WMAT -DSBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

## Database :

Published Applications NA:\*

- 1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/prodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 5: /cgn2\_6/prodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 6: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 7: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 8: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 9: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 10: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 11: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 13: /cgn2\_6/prodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/prodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 15: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 17: /cgn2\_6/prodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/prodata/1/pubpna/US10F\_PUBCOMB.seq.\*
- 19: /cgn2\_6/prodata/1/pubpna/US10G\_PUBCOMB.seq.\*
- 20: /cgn2\_6/prodata/1/pubpna/US10H\_PUBCOMB.seq.\*
- 21: /cgn2\_6/prodata/1/pubpna/US10I\_PUBCOMB.seq.\*
- 22: /cgn2\_6/prodata/1/pubpna/US10J\_PUBCOMB.seq.\*
- 23: /cgn2\_6/prodata/1/pubpna/US10K\_PUBCOMB.seq.\*
- 24: /cgn2\_6/prodata/1/pubpna/US10L\_PUBCOMB.seq.\*
- 25: /cgn2\_6/prodata/1/pubpna/US10M\_PUBCOMB.seq.\*
- 26: /cgn2\_6/prodata/1/pubpna/US10N\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1444	100.0	80161	17	US-10-329-148A-1
2	441.5	30.6	849	15	US-10-156-761-925
3	441.5	30.6	9025608	15	US-10-156-761-1
4	438.5	30.4	846	16	US-10-132-134-21
5	438.5	30.4	52101	16	US-10-132-134-1
6	435.5	30.2	88400	17	US-10-329-148A-1
7	412	26.1	852	21	US-10-844-716-1
8	376.5	26.1	852	10	US-09-953-348-36
9	376.5	26.1	852	15	US-10-267-255-36
10	376.5	26.1	53500	10	US-09-953-348-76
11	376.5	26.1	53500	15	US-10-267-255-76
12	361.5	25.0	30000	11	US-09-980-217-1
13	345	23.9	828	10	US-09-953-348-35
14	345	23.9	828	15	US-10-267-255-35
15	268	18.6	52659	21	US-10-496-377-1
16	250	17.3	843	16	US-10-279-029-54
17	250	17.3	843	16	US-10-219-810-17
18	250	17.3	858	17	US-10-369-493-42625
19	246.5	17.1	843	16	US-10-279-029-53
20	246.5	17.1	843	16	US-10-219-810-16
21	242.5	16.8	1041	19	US-10-257-344A-13
22	241.5	16.7	1444	10	US-09-885-723-21
23	241.5	16.7	1444	21	US-10-862-907-21
24	240	16.6	930	16	US-10-219-810-44
25	238.5	16.5	1386	18	US-10-425-114-20190
26	238.5	16.5	1582	18	US-10-425-114-3622
27	238.5	16.5	1595	18	US-10-425-114-28071
28	238	16.5	1370	18	US-10-425-114-14934
29	238	16.5	1381	18	US-10-425-114-14931
30	237	16.4	1551	18	US-10-424-559-54894
31	236.5	16.4	1477	9	US-09-779-144A-6
32	234.5	16.2	1230	19	US-10-437-963-65029
33	232.5	16.1	1089	19	US-10-437-963-45285
34	232.5	16.1	1630	19	US-10-437-963-45286
35	232	16.1	1044	16	US-10-219-810-8
36	232	16.1	2590	17	US-10-418-555-151
37	229.5	15.9	681	21	US-10-487-901-2303
38	229.5	15.9	681	21	US-10-487-901-6303
39	229.5	15.9	1038	16	US-10-279-029-43
40	229.5	15.9	1038	16	US-10-219-810-4
41	229.5	15.9	1038	19	US-10-668-240-1
42	229.5	15.9	7713	19	US-10-668-240-5
43	229.5	15.9	8179	19	US-10-668-240-5
44	228	15.8	1044	16	US-10-219-810-9
45	226	15.7	1131	16	US-10-279-029-44

## ALIGNMENTS

RESULT 1  
US-10-329-148A-1  
; Sequence 1, Application US/10329148A  
; Publication No. US20040023343A1  
GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H  
; APPLICANT: Broughton, Mary C  
; APPLICANT: Crawford, Kathryn P  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Treadway, Patti J  
; APPLICANT: Turner, Jan R  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: BioSynthetic Genes For Spinosyn Insecticide  
; FILE REFERENCE: 50489 DIV1  
; CURRENT APPLICATION NUMBER: US/10/329,148A  
; CURRENT FILING DATE: 2002-12-23

```
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/603,207B
/ PRIOR FILING DATE: EARLIER FILING DATE: 2000-06-23
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/370,700
/ PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-09
/ NUMBER OF SEQ ID NOS: 39
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 1
/ LENGTH: 80161
/ TYPE: DNA
/ ORGANISM: Saccharopolyspora spinosa
US-10-329-148A-1

Alignment Scores:
Pred. No.: 1,29e-160 Length: 80161
Score: 1444.00 Matches: 275
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-069-353A-8 (1-275) x US-10-329-148A-1 (1-80161)

QY 1 ValLeuProGlyValAlaProThrsSerGlnGlnValGlyGlnMetTyraSpleuValThr 20
DB 20168 GTGTGGCCAGTGGCCGACCAACATCGCAGAGTTGGGCAATGATGACCTGGTCAGC 20227
QY 21 ProLeuLeuAsnSerValAlaGlyProCySaAlaIleHisGlyTyrTrpGluAsn 40
DB 20228 CCGTGTCTAACTCGGTCCGGGGCGCCCTGTGGCCATCCACACGCTACCTGGGAGAAC 20287
QY 41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
DB 20288 GACGGCGCGGCTTCCTGGACAGAGCGCGCGACCGGCTACACGACCTTGTCCCGAAGC 20347
QY 61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCyGlyThrGlyGlnProAla 80
DB 20348 ACCGGTCTGATGGCGCGCTGACCTGATGTGGGGTGGCGTACCGACACACGAGC 20407
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
DB 20408 CTGGCGGTGGCGCGACACCGGATCCGATCACCGGATCACCGCTCACCGACGGTGCNA 20467
QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
DB 20468 GTGGCCATGCGCGCTGATGGCGACGCAACCGGACCTAAGCCACCGGTTGACTTCTCG 20527
QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
DB 20528 TCCGTCGATGCGATGTCCCTGCGGTAACCGGACAAATGCTTTCGACGCCCTGGGCGCATG 20587
QY 141 GlnSerLeuLeuGluMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160
DB 20588 CAGTGCCTTTGAGATGTCCCAACCGGACCGGCTGATCCGGGAAATCTTTCGAGTACTC 20647
QY 161 LysProGlyGlyIleLeuGlyValThrGluValIleLysArgGluAlaGlyGlyMet 180
DB 20648 AAACCCGGTGCATCTCGCGCTCACCGAGTCTGTAACGAGAAAGCGGCGCGGAGTGG 20707
QY 181 ProValSerGlyAspArgTrpProThrGlyLeuAlaGlyIleCysLeuAlaGlnLeuLeu 200
DB 20708 CCGGTGTCCGGGAGACGAGTGGCGACCGGCTTCGATGTGCTGCTGAGCAACTTCTCG 20767
QY 201 GlnSerLeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220
DB 20768 GAATGCTCTCGTCAGCGGGGTTCCAGATCTCGATTGGAGAGAGCTGTCTCGTGGAGACC 20827
QY 221 ArgTyrPheMetProGlnPheAlaGlnGluLeuAlaAlaHisGlnHisGlyIleAlaAsp 240
DB 20828 CCGTACTTCATGCGCGAGTTCGCCGAAGAGCTCGCTGGCCACACAGACGGAGTCCGGAGC 20887
QY 241 ArgTyrGlyProAlaValAlaGlyTrpAlaAlaAlaValCysAspTyrGlyLysTyrAla 260
DB 20888 AGGTACGGGCGGCTGTCTGCGCGGCTGGCGCGGCTGTGCAATTATGAGAAATATATGCC 20947
```

```
QY 261 HisAspMetGlyTyrAlaIleLeuThrAlaArgLysProValGly 275
DB 20948 CACGACATGGCTATGCAATTCCTGACGGCGGAGGCGGTCGGC 20992

RESULT 2
US-10-156-761-925
/ Sequence 925, Application US/10156761
/ Publication No. US20030119018A1
/ GENERAL INFORMATION:
/ APPLICANT: OMURA, SATOSHI
/ APPLICANT: IKEDA, HARUO
/ APPLICANT: ISHIKAWA, JUN
/ APPLICANT: HORIKAWA, HIROSHI
/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
/ APPLICANT: HATTORI, MASAHIRA
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-262
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ PRIOR FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ PRIOR FILING DATE: 2001-08-02
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 925
/ LENGTH: 849
/ TYPE: DNA
/ ORGANISM: Streptomyces avermitilis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(849)
US-10-156-761-925

Alignment Scores:
Pred. No.: 7,65e-43 Length: 849
Score: 441.50 Matches: 111
Percent Similarity: 56.32% Conservative: 45
Best Local Similarity: 40.07% Mismatches: 107
Query Match: 30.57% Indels: 15
DB: 15 Gaps: 6

US-10-069-353A-8 (1-275) x US-10-156-761-925 (1-849)

QY 7 ProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThrProLeuLeuAsnSerVal 26
DB 28 CCTCTTCCCTGGAAGTGGGGACTTACGACCGGTTGACCGACCTCATGAAATCGTGGC 87
QY 27 AlaGlyGlyProCysAlaIleHisGlyTyrTrpGluAsnAspGlyArgAlaSerTrp 46
DB 88 CTGGGTGG-----MACACCACTCGGATATCTGGCCGCGACCGGGGAGCGGACATTCA 141
QY 47 Gln---GlnAlaAlaAspArgLeuThrAspLeuValAlaGluArgThrValLeuAspGly 65
DB 142 CCGGGCAAGCGCGCGACCGGCTCACCGATCTCTCATCGGCATCGGAGAGGATCAGC 201
QY 66 GlyValAlaGluLeuAspValGlyCyGlyThrGlyGlnProAlaLeuArgValAlaArg 85
DB 202 GCGCGCGGCTCGGAGAGTGGCTGGGTTCCGAAAGCGGAGCGGCGGCTGGCCCTG 261
QY 86 AspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValAlaIleAla 105
DB 262 AGCGCGCCCTCGATGTCTGGCGCTGACGCTGAGCGGATTCAGTTCGGGCTGGCCAGC 321
QY 106 AspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSerCysValAspAlaMet 125
DB 322 GCTCTCGGAGAGCGTGCAGCTGAGCGGACCGGATCGGTTCACCGCGCGAGCGGATG 381
QY 126 SerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMetGlnSerLeuLeuGlu 145
DB 382 GAACGTGCGTTTCCCGACGGGCTCTTGCACCGGCGGTGGGCGCTCGAGTGTCTCTGAC 441
```

```
QY 146 MetSerGluProAspArgAlaIleArgGluIleLeuArgValLeuLysProGlyGlyIle 165
DB 442 ATGCCAGACCCCGCACAGGATGATCGCGGAGATCGCCGGGTCTCGCCCGCCCGCCGCG 501
QY 166 LeuGlyValThrGluValValLysArg-GluAlaGlyGlyLysMetProValSerGlyLys 185
DB 502 CTGGCCGTCACGGACGTCGACCTGCGCGCTTGGGGGAGCCGGCATGAAAGCCGGGGAG 561
QY 185 pArgTrpProthnGlyLeuArgIle-----CysLeuAlaGluGlnLeuGluIse 202
DB 562 TGACACTCCACAGTGTGCTCGCGGTCGCCGCCCTGTGCACATGACAGATCGCCGGAATG 621
QY 202 rLeuArgAlaAlaGlyPheGluIleLeuSerTrpGluAspValSerSerArgThrArgTrg 222
DB 622 -ATCGCGGACGCGCGCGCTGAACTGACATGACGACGACGATCGGCGATCAGGTC----- 675
QY 222 rPheMetProGlnPheAla-----GluGluLeuAlaAlaIleGlnIleGlyIleAl 239
DB 676 -GTGGGCCCCCTTTTGGCCGCGCTGTCACAGTGAACAGACACTCGACAGATACGC 734
QY 239 aAspArgTrgGlyProAlaValAlaGlyTrp-----AlaAlaAlaValCysAspTrgAl 257
DB 735 GGGCGGCTTCGGAGATGGCGGTGCGGAGATGCGGAAGTGTTGCACAGTGCACAGCGCT 794
QY 257 uLysTrgAlaIleAspMetGlyTrgAlaIleLeuThrAlaArgLysPro 273
DB 795 CCCCTGGACGCGGACATCGCTATGTCTGTGACGCGCGCGCGCGCG 843

RESULT 3
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OR INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 1,16e-37 Length: 9025608
Score: 441.50 Matches: 111
Percent Similarity: 56.32% Conservative: 45
Best Local Similarity: 40.07% Mismatches: 107
Query Match: 30.57% Indels: 15
DB: 15 Gaps: 6

US-10-069-353a-8 (1-275) x US-10-156-761-1 (1-9025608)
QY 7 ProThrSerGlnGlnValGlyGlnMetTrpAspLeuValThrProLeuLeuAsnSerVal 26
DB 1137453 CCCCTTCCCTCGAAGATGGGGAGCTACTACGACCGTTTGAACGACCTCATGAACTCGGCG 1137394
```

```
QY 27 AlaGlyGlyProCysAlaAlaIleHisIleGlyTrgTrpGluAsnAspGlyArgAlaSerTrp 46
DB 1137393 CTGGGGTGG-----AACACCACTCTCGATCTGCGCCGACCGGGGAGCGGATCTCA 1137340
QY 47 Gln---GlnAlaAlaAspArgLeuThrAspLeuValAlaGluArgThrValLeuAspGly 65
DB 1137339 CCGGGAGAGCGCGCGACCGGCTCAACCGATCTCTCATCGCAGAGCTGAGAGGATCAACG 1137280
QY 66 GlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAlaLeuArgValAlaArg 85
DB 1137279 GGCCTGCGGCTCTGACCTCGCTGCGCTGCGGTTCCGGAACCGCGCGCTGCGCTGCG 1137220
QY 86 AspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGlnValAlaIleAlaAla 105
DB 1137219 AGCGGCGCGCTGAGTGTCTGCGCTGACGAGTGAAGGATTCAGGTCCGGCTGGGAC 1137160
QY 106 AspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSerCysValAspAlaMet 125
DB 1137159 GCTCTGGAAGAGATGCGACGATGGCGGACCGGGTCTGTTCACCCGTCGACGCGATG 1137100
QY 126 SerLeuProTrpProAspAsnAlaPheAspAlaAlaTrpAlaMetGlnSerLeuGlu 145
DB 1137099 GAATGCTCGTTTCCGACGCGGTCTTCACGCGGCTGGGCTGGAGTGTCTCTGCAC 1137040
QY 146 MetSerGluProAspArgAlaIleArgGluIleLeuArgValLeuLysProGlyGlyIle 165
DB 1137039 ATGCCAGCCCGCACAGGATGATCGGGAGATCGCCGGGTCTCGCCCGCGCGCGCG 1136980
QY 166 LeuGlyValThrGluValValLysArg-GluAlaGlyGlyLysMetProValSerGlyLys 185
DB 1136979 CTGGCGGTCAAGAGATGTCACATGCGCGCTTGGGGGAGACCGGATGAAAGCGCGGAG 1136920
QY 185 pArgTrpProthnGlyLeuArgIle-----CysLeuAlaGluGlnLeuGluIse 202
DB 1136919 TGCACGCTCCAGTGTCTCGCGGTCGCCGCCCTGTGTCACATGACAGATACCGCGGAATG 1136860
QY 202 rLeuArgAlaAlaGlyPheGluIleLeuSerTrpGluAspValSerSerArgThrArgTrg 222
DB 1136859 -ATGCCGACGCGCGGCTGAACTGACATGACGACCGACATCGGGAGTCAAGTCTC----- 1136806
QY 222 rPheMetProGlnPheAla-----GluGluLeuAlaAlaIleGlnIleGlyIleAl 239
DB 1136805 -GTGGGCCCCCTTTTGGCGCGCTGTCGTCGACAGTGAACAGACACTCGACAGATACGC 1136747
QY 239 aAspArgTrgGlyProAlaValAlaGlyTrp-----AlaAlaAlaValCysAspTrgAl 257
DB 1136746 GCGCGGCTTCGGAGATCGCGGTGCGGAGATGCGGAAGTGTTGCACAGTGCACAGCGCT 1136687
QY 257 uLysTrgAlaIleAspMetGlyTrgAlaIleLeuThrAlaArgLysPro 273
DB 1136686 CCCCTGGACGCGGACATCGCTATGTCTGTGACCGCGCGCGCGCG 1136638

RESULT 4
US-10-132-134-21
; Sequence 21, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffe, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OR INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-205
; CURRENT APPLICATION NUMBER: US/10/132,134
; PRIOR FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 846
; TYPE: DNA
; ORGANISM: Streptomyces platensis subsp. roseaceus
US-10-132-134-21
```

## Alignment Scores:

Pred. No.: 1,75e-42 Length: 846  
 Score: 438.50 Matches: 106  
 Percent Similarity: 56.10% Conservative: 55  
 Best Local Similarity: 36.93% Mismatches: 95  
 Query Match: 30.37% Indels: 31  
 DB: 16 Gaps: 9

US-10-069-353A-8 (1-275) x US-10-132-134-21 (1-846)

```

QY      2 LeuProGlyGlyAlaProThSerGlnGlnValGlyGlnMetCysThrAspLeuValThrPro 21
DB      34 GTCCCGCTCCGCCCCCGCTCCCGAAAGGTGGACACCTCTACACCGCTTACCGCGCA 93
QY      22 LeuLeuAsnSerValAlaGlyGlyProCysAlaIleHisGlyTyrThrGlu--Asn 40
DB      94 CTGGACACCGAAGCGCGCGCGCG-----AGCTTCACCTCGGCTACTGGAGCTTCGAC 147
QY      41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
DB      148 GACAAACGACACCGCGCTCGTGGAAAGCGCGCGACCGGCTCACCGACACGATGACCGACCGC 207
QY      61 ThrValLeuAspGlyGlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80
DB      208 CTGGCGATGACACGAGGACAGCGCGGCTCTGACGCTGCGGTGGAGTCCGACGCGCGCGC 267
QY      81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
DB      268 ATGCGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 327
QY      101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
DB      328 ATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 387
QY      121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaIleAspAlaAlaThrAlaMet 140
DB      388 CACGCGCGACGATGAACTGACCTGCTCCCGACGACTCTTGACGCGCGCGCGCGCGCGCGCG 447
QY      141 GlnSerLeuLeuGlnIleMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160
DB      448 GAGTGCATCTTCCACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 507
QY      161 LysProGlyGlyIleLeuGlyValThrGluValValLysArgGluAlaGlyGlyMet 180
DB      508 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 552
QY      181 ProValSerGlyAspArgTrpProThrGly-----LeuArg 192
DB      553 CCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 612
QY      193 IleCysLeuAlaGlnIleLeuGlnIleSerLeuArgAlaAlaGlyPhe-----Glu 209
DB      613 CTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 672
QY      210 IleLeuAspTrpGluAspValSerSerArgThrArgTyr--PheMetProGlnPheAla 228
DB      673 CTCTCTGACATGACCGGACGAGCGGTGCTGACGCTTCCGACGACGATGAGCGAGGCTTCC 732
QY      229 GlnGluLeu--AlaAlaHisGlnHisGlyIleAlaAspArgTyrGlyProAlaValAla 247
DB      733 CAGGAGATGACGACCGCTTTCGACGACGAGGACGAGGACGAGGACGAGGACGAGGACGAGG 783
QY      248 GlyTrpAlaAlaAlaValCysAspTrpGlyGlyValAlaHisIleAspMetGlyTyrAlaIle 267
DB      784 -----GCCTCATGATGACGTGACGAAATTC-----GCCTCGGTTCTG 822
QY      268 LeuThrAlaArgLysProVal 274
DB      823 CTGACCGCGCGCAAAAGCGCTCTC 843

```

RESULT 5  
 US-10-132-134-1  
 ; Sequence 1, Application US/10132134

```

; Publication No. US20030171562A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Zafopoulos, Emmanouel
; APPLICANT: Stalfi, Alfredo
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-2US
; CURRENT APPLICATION NUMBER: US/10/132,134
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 52101
; TYPE: DNA
; ORGANISM: Streptomyces platensis subsp. rosaceus
; US-10-132-134-1

```

## Alignment Scores:

Pred. No.: 3.5e-40 Length: 52101  
 Score: 438.50 Matches: 106  
 Percent Similarity: 56.10% Conservative: 55  
 Best Local Similarity: 36.93% Mismatches: 95  
 Query Match: 30.37% Indels: 31  
 DB: 16 Gaps: 9

US-10-069-353A-8 (1-275) x US-10-132-134-1 (1-52101)

```

QY      2 LeuProGlyGlyAlaProThSerGlnGlnValGlyGlnMetCysThrAspLeuValThrPro 21
DB      51267 GTCCCGCTCCGCCCCCGCTCCCGAAAGGTGGACACCTCTACACCGCTTACCGCGCA 51326
QY      22 LeuLeuAsnSerValAlaGlyGlyProCysAlaIleHisGlyTyrThrGlu--Asn 40
DB      51327 CTGGACACCGAAGCGCGCGCGCG-----AGCTTCACCTCGGCTACTGGAGCTTCGAC 51380
QY      41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
DB      51381 GACAAACGACACCGCGCTCGTGGAAAGCGCGCGCGCGCTCACCGACACGATGACCGACCGC 51440
QY      61 ThrValLeuAspGlyGlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80
DB      51441 CTGGCGATGACACGAGGACAGCGCGGATCTCTGACGCTGCGGTGGAGTCCGACCGCGCGC 51500
QY      81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
DB      51501 ATGCGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 51560
QY      101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
DB      51561 ATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 51620
QY      121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaIleAspAlaAlaThrAlaMet 140
DB      51621 CACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 51680
QY      141 GlnSerLeuLeuGlnIleMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160
DB      51681 GAGTGCATCTTCCACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 51740
QY      161 LysProGlyGlyIleLeuGlyValThrGluValValLysArgGluAlaGlyGlyMet 180
DB      51741 CGCGCGCGCGCGCGCGCGCTGCTCTCACGACCTTTCGACGAGCGCGCG-----LeuArg 192
QY      181 ProValSerGlyAspArgTrpProThrGly-----LeuArg 192
DB      51786 CCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 51845
QY      193 IleCysLeuAlaGlnIleLeuGlnIleSerLeuArgAlaAlaGlyPhe-----Glu 209
DB      51846 CTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 51905
QY      210 IleLeuAspTrpGluAspValSerSerArgThrArgTyr--PheMetProGlnPheAla 228

```

```

Db      51906 CTCCTGACATCCAGGAGAGCGTGCCTGAGACCTTGAGCAGATGAGCCAGGCTCC 51965
      229 GJUGLuleu---AlaAlaHISglnHISgLYleAlaAspArgTYrgLYProAlaValAla 247
      51966 CAGGAGATGCAGACCGCTTTCAGACGACGAGGAGAGAAAGTTTCAGCCCC----- 52016
Qy      248 GJYTPAlaAlaAlaValaLYsaAspTYrgLYuLYsTYrAlaHISAspMetGLYTYrAlaIle 267
      52017 -----GCCCTCATGATCGACGTGACGAAATTC-----GGCTCCGTTCTG 52055
Qy      268 LeuThrAlaArgLYsProVal 274
      52056 CTGACCGCCCAAAAGCCCTC 52076

RESULT 6
US-10-329-148A-1/c
; Sequence 1, Application US/10329148A
; Publication No. US20040023343A1
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madhuri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/10/329,148A
; PRIOR FILING DATE: 2002-12-23
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/603,207B
; PRIOR FILING DATE: EARLIER FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/370,700
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
US-10-329-148A-1

Alignment Scores:
Pred. No.: 1,4e-39 Length: 80161
Score: 435.50 Matches: 108
Percent Similarity: 51.67% Conservative: 47
Best Local Similarity: 36.00% Mismatches: 96
Query Match: 30.16% Indels: 49
DB: Gaps: 11

US-10-069-353A-8 (1-275) x US-10-329-148A-1 (1-80161)
Qy      7 ProThSerGlnGlnValGlyGlnMetTYrAspLeuValThrProLeuLeuAAsnSerVal 26
      13562 CCGACCCCGCATCGAGGAGTGCATCTTCATCCGTTG-----GCG 13521
Db      27 AlaGLYGLYProCYsaAlaIleHISgLYTYrTP-----GluAAsnAsp 41
      13520 CACGGGGCTGCC-----CTGCACACCGTTACTGCGGGCGGGGTATCGGAGAGATGCC 13467
Qy      42 GJYrAspAspSerTPGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArgThr 61
      13466 GGTGCCACACCGTGTGTGGATGTGCGACCAATGACCGACTTCATCGACCAAGGCC 13407
Db      62 ValLeuAspGLYGLYValaArgLeuAspValGLYCYsGLYTHrGlyGlnProAlaLeu 81
      13406 GCGCTCGTCGCCGCGACCACTGTGTGACCTGGAGCTGGCGCAATGGAGCCCGTATGTC 13347
Qy      82 ArgValAlaArgAspAspAlaIleGlnIleThrGlyIleThrValSerGlnValGlnVal 101
      13346 GTTCCGGCATGCGCCAGCGGGCTTCGAGTCCCGGAATCACCGGACCCAGCATCTC 13287

```

```

Qy      102 AlaIleAlaAlaAspCYsaValaArgGluArgGlyLeuSerHISArgValaAspPheSerCys 121
      13286 GCGCGCCGACCAAGCTGCCAAGACCGGACTGGCCGGAGCTTGAGTTGATGCTA 13227
Qy      122 ValaAspAlaMetSerLeuProTYrProAspAspAlaPheAspAlaIleTPAlaMetGln 141
      13226 GTCGACGGCGCCCACTGCGCTTACCAGCGGTTCTTCAAGCCGCGCATGGCGCATGACG 13167
Qy      142 SerLeuLeuGluMetSerGlnProAspArgAlaIleArgGlnIleLeuArgValLeuArg 161
      13166 TCCGTCTGTCAGATCGTGCAGACGAGCCCGCGGATCCGAGAGTCCACCGAAATCTTGAA 13107
Qy      162 ProGlyGLY-----IleuGLY-----ValThrGluValaValaArgGluAlaGly 177
      13106 CCGGGCGCGGTTGCTCTCGGAGACATCATCATCTCGGTTTCGATCCCGGAGAGTAC 13047
Qy      178 GJYGLYMetProValSerGLYAspArgTYrProThrGlyLeuArgIleCYsaLeuAlaGlu 197
      13046 GCGGGCGGT-----TGG-----ACGGGACGACGCGCCCATCTTGAAAC 13008
Qy      198 GlnLeuLeuGluSerLeuArgAlaAlaGlyPheGlnIleLeuAspTYrGluAspValSer 217
      13007 AGCTTCACCGCGCTGTGCAGACCGCGGTTCTGAGATTCTGAAGTCAACGACCTCAGC 12948
Db      218 SerArgTYrArgTYrPheMetProGlnPheAlaGluGluLeu----- 231
      12947 GCACAGACCAAGTGCATGTGTTCTCTGTCGTGACGAGATTCTCCGGAACCTCATGAG 12888
Qy      232 -----AlaAlaHISglnHIS-----GlyIle 238
      12887 CTCGCGCGGCTGAGCGCTGCGGTGCGGACCTTACACCAACGTACTTGGAGACATC 12828
Qy      239 AlaAspArgTYrGLYProAlaValaIleGlyTPAlaAlaAlaValCYsaAspTYrGluLYs 258
      12827 GCGGCGACAGCAGGACCGGACACAGCATGATCGCGGTTGCGGAATPCCGGA 12768
Db      259 -----TYrAlaHISAsp-----MetGLYTYrAlaIleLeuThrAlaArgLYs 272
      12767 CATCCGATTCAGCCAGAAACAGAGAAAGCATGGTTTCATGCTCTCGAGGCTGAAAG 12708

RESULT 7
US-10-844-716-1/c
; Sequence 1, Application US/10844716
; Publication No. US20050003409A1
; GENERAL INFORMATION:
; APPLICANT: Huang, Chenglin
; APPLICANT: Chaleff, Deborah T.
; APPLICANT: Ruppen, Mark B.
; APPLICANT: Stephens, Jerome
; TITLE OF INVENTION: Cloning Genes From Streptomyces Cyaneogriseus Subsp.
; FILE REFERENCE: AM100484
; CURRENT APPLICATION NUMBER: US/10/844,716
; CURRENT FILING DATE: 2004-05-13
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 88400
; TYPE: DNA
; ORGANISM: bacteria
US-10-844-716-1

Alignment Scores:
Pred. No.: 1.05e-36 Length: 88400
Score: 412.00 Matches: 103
Percent Similarity: 50.88% Conservative: 42
Best Local Similarity: 36.14% Mismatches: 106
Query Match: 28.53% Indels: 34
DB: Gaps: 9

US-10-069-353A-8 (1-275) x US-10-844-716-1 (1-88400)
Qy      5 GJYAlaProThrSerGlnGlnValGlyGlnMetTYrAspLeuValThrProLeuLeuAAsn 24

```

```

Db      51282 GGATATACCCAGCCCGCGAGCTGGTGACTACTACGACCGCATGACCCGCTTGAAC 51223
Qy      25 SerValAlaGlyIleProCysAlaIleHisGlyIleThrGluAsn---AspGlyArg 43
Db      51222 CGGGCCCTGGGGGGC-----AACAGCAGCTCTGGATCTGGCCGACCGGACGAGCG 51169
Qy      44 AlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArgThrValLeu 63
Db      51168 AACACACTCGGCAGAGCTCTCCAGCGCTCACCGACCATATGATCGGCAAGCTTCGGGA 51109
Qy      64 AspGlyGly-----ValArgLeuLeuAspValGlyCysGlyIleThrGlyIleProAla 80
Db      51108 CAACCCGGGCGCCCGGTACAGAGTGTGACGTGACGTGCGCTGCGCGCGCGCGCGCG 51049
Qy      81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
Db      51048 CTGGCCCTGGCGCAGCGACCGCGGTGACATGTCGGGATCACCATTCAGCCCCCGGCGAG 50989
Qy      101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
Db      50988 GTTCAGACTGGCCACCGCGCTCGCCGAGCGGTCCGAGCTCGGACCGGGTCCGCTTCAG 50929
Qy      121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
Db      50928 TCCGCCGAGCGCATGAGCTGCGCTTCCCGACGCGCTTCGACGCGCTCGGCGCGCTTC 50869
Qy      141 GlnSerLeuLeuGlnMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160
Db      50868 GAGTCTCTCTGACATGCTCCGACCGCGCGCGGTTCAGAGATGAGCCCGTGTGCTG 50809
Qy      161 LysProGlyGlyIleLeuGlyValThrGluValValLysArgGluAlaGlyIleGlyMet 180
Db      50808 CCCCCCGGTGGCGGTGGCGGCGCATGACGTACACATCCGCGCATCA----- 50761
Qy      181 ProValSerGlyAspArgTrpProThrGlyLeuAlaGlyIleCysLeuAlaGlnLeuLeu 200
Db      50760 -----CAGCCGAGGGGGCGGACTGTGTCTCCAGCGCACTCCCGCTC 50719
Qy      201 GlnSerLeu-----ArgAlaAlaGlyPheGluIle 210
Db      50718 CCCTGCTGATCCCATCATCCGCTATGCGGGATGATAGAGGAGCGCGGACTCCGGCTG 50659
Qy      211 LeuAspTrpGluAspValSerSerArgThrArgTyrPheMetProGlnPheAla----- 228
Db      50658 ACCGAGCTCACCGATCGGTAGACACGTC-----ATCGCCCGTCGTAACGCGCATG 50605
Qy      229 ---GluGluLeuAlaAlaHisGlnHisGlyIleAlaAspArgTyrGlyProAlaValAla 247
Db      50604 GGTGACGAGTACGGCGCAACCGCCCTACGCGGAGCCCTGAGATGACGCGCGAGC 50545
Qy      248 GlyTrpAlaAlaAlaVal-----CysAspTyrGluLysTyrAlaHisAspMetGlyTyr 265
Db      50544 GACCTGGAGACCTTCTGTGGGCAAGTGCAGC---CACTGTACACGAGAGACATCGGGTAC 50488
Qy      266 AlaIleLeuThrAla 270
Db      50487 GTGTCCTGACCGCG 50473

```

```

/ PRIOR APPLICATION NUMBER: PCT/US00/06394
/ PRIOR FILING DATE: 2000-03-10
/ PRIOR APPLICATION NUMBER: 09/266965
/ PRIOR FILING DATE: 1999-03-12
/ NUMBER OF SEQ ID NOS: 153
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 36
/ LENGTH: 852
/ TYPE: DNA
/ ORGANISM: Streptomyces lavendulae
US-09-953-348-36

Alignment Scores:
Pred. No.: 4,86e-35 Length: 852
Score: 376.50 Matches: 101
Percent Similarity: 50.84% Conservative: 51
Best Local Similarity: 33.78% Mismatches: 94
Query Match: 26.07% Indels: 53
DB: 10 Gaps: 11

US-10-069-353A-8 (1-275) x US-09-953-348-36 (1-852)

Qy      2 LeuProGlyGlyAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThrPro 21
Db      25 CTCCCATGCTCCCTCACCTCGCTCGAGGAGATGGGCGGCTCTACGACCGGTTGACCGCG 84
Qy      22 LeuLeuAspSerValAlaGlyIleProCysAlaIleHisGlyIleThrGluAsn--- 40
Db      85 CTGGAGAGCGGCTCCCTCGGAG-----AACTGCACTTCGGCTACTGGGACTCCCCC 138
Qy      41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
Db      139 GACAGCCAGGTGGCGCTGCGCGGAGGCCACCGCGGTCTACCGACATGATGCGGAGCGG 198
Qy      61 ThrValLeuAspGlyValArgLeuLeuAspValGlyCysGlyIleThrGlyIleProAla 80
Db      199 CTGGCATGCGCGCGGCTCCCGCTCGGTCTGAGCTCGGTGCGCGGTGGGACCCCGGCG 258
Qy      81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
Db      259 GTACCATGCTCCCGCTCAGCGAGCGGATGATCAAGGCACTCTGGTGAAGCATGAGAGAG 318
Qy      101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
Db      319 GTGTCGCGGCGCAACGCGCTGCGGAGGAGCGGGGCTCGCGGACCGGGCGCTTCAG 378
Qy      121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
Db      379 CGGGCCGAGGATGAGACTCTCCCTTGAGGACGAGCTTCAAGCGCGCTCATCGCTTC 438
Qy      141 GlnSerLeuLeuGlnMetSerGluProAspArgAla-----IleArgGluIleLeuArg 158
Db      439 GAATGATCATCCACATG-----CCGACCGGCGCCAGGTCGTGCGCCAGGTGGCGGG 492
Qy      159 ValLeuLysProGlyGlyIleLeuGlyValThrGluValValLysArg----- 174
Db      493 GTGCTGCGGCGCGGGGCGGTCTGTGCTCACCGACTTCTTCGAGCGGGCCCCCTGCGCC 552
Qy      175 ---GluAlaGlyIleGlyMetPro-----ValSer 183
Db      553 CCCGAGGGGCGGGCGCGCTCAGCGCTACTCCACGACTTCATGATGACCATGAGTACG 612
Qy      184 GlyAspArgTrpProThrGlyLeuArg-----IleCys 194
Db      613 GCCAGGGGTACCTCCCTGCTCGGGGGCGGGCGCTGCGGTGAGAGATTCTCTGAC 672
Qy      195 LeuAlaGlnGlnLeuLeuGlnSerLeuArgAlaAlaGlyPheGluIleLeuAspTrpGlu 214
Db      673 ATCAGCGACGACCGCTGAGAGAGC-----TTCAGAGCTGCTTCG---GAG 717
Qy      215 AspValSerSerArgThrArgTyrPheMetProGlnPheAlaGlnGluLeuAlaHis 234
Db      718 CGCATCACTCTCGAAGAGAGGCTGAGAGCGAGATTGGCGAGAGATG---GTGAAAC 774

```

```

RESULT 8
US-09-953-348-36
/ Sequence 36, Application US/09953348
/ Publication No. US20030134398A1
/ GENERAL INFORMATION:
/ APPLICANT: Sherman, David. H
/ APPLICANT: Mao, Yinying
/ APPLICANT: Varoglu, Mustafa
/ APPLICANT: He, Min
/ APPLICANT: Sheldon, Paul
/ TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER
/ FILE REFERENCE: 600.530US1
/ CURRENT APPLICATION NUMBER: US/09/953,348
/ CURRENT FILING DATE: 2001-09-12

```





```

Db      7560 CTCGCCATGCGCCCTCACCCGCTCCGAGAGGCGGCGCTACGACCGGTTCCACCGCG 7501
      22 LeuLeuAspSerValAlaIleGlyProCysAlaIleHisIleGlyThrTrpGluAsn--- 40
      7500 CTGGAGCGCGGCTCTCTCCGCGAG-----AACCTGCATCTTGCGGCTACTGGAGCTCCGCC 7447
      41 AspGlyArgAlaSerTrpGlnGlnAlaIleAspArgLeuThrAspLeuValAlaGluArg 60
      7446 GACAGCGAGTGCGCGCTGCGCGAGGCGACCGCGCTCACCAGCATGATGGCCGAGCGG 7387
      61 ThrValLeuAspGlyGlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80
      7386 CTGGCGATGCGCGCGCTCCGCGCTCCGAGCATTCGCGCTGCGCGCGCGCGCGCGCG 7327
      81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
      7326 GTACCGATCGCGCGCTCCGAGCGGCGCATGTCAAGGACATTCGGTAGCCATGAGCAG 7267
      101 ValAlaIleAlaIleAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
      7266 GTGCTCGCGGCGCAAGCGCTGCGCGAGGCGCGGCTGCCGACCGCGCGGCTTCCAG 7207
      121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
      7206 CCGCGCGAGCGCATGAGCTCCCTTCGAGAGCGAGAGCTTCGAGCGCTCATCGCCCTC 7147
      141 GlnSerLeuLeuGlnMetSerGlnProAspArgAla-----IleArgGlnIleLeuArg 158
      7146 GAATGATATCATCAATG-----CCGAGCGCGCGCGCGAGCTCGCCAGGCTCGCGCG 7093
      159 ValLeuArgProGlyGlyIleLeuGlyValThrGluValAlaValAspArg----- 174
      7092 GTGCTCGCGCGCGCGCGCGCTGCTGCTGCTCAGCATTCCTTGAGGCGCGCGCGCGCG 7033
      175 ---GluAlaGlyGlyIleMetPro-----ValSer 183
      7032 CCGAGGCGCGCGCGCGCGCTCCAGCGCTACCTCCAGCATTCATGATGACCATGTGTACG 6973
      184 GlyAspArgTrpProThrGlyLeuArg-----IleCys 194
      6972 GCGGAGCGGTACCTCCCTCGCTGCGGCGGCGCGCGCTGCGTGGAGGAGTTCTCCGAC 6913
      195 LeuAlaGlnGlnLeuLeuGlnSerLeuArgAlaAlaGlyPheGlnIleLeuAspTrpGln 214
      6912 ATCAGCAGCACGACCTCGAGAAAGACC-----TTCAGGCTGCTTCG---GAG 6868
      215 AspValSerSerArgThrArgTyrPheMetProGlnPheAlaGlnGluLeuAlaHis 234
      6867 CGCATCACTCTCGAAGCAGAGGCTGAGAGCGAGTTCTGCGAGAGATG---GTGAC 6811
      235 GlnHisGlyIleAlaAspArgTyrGlyProAlaValAlaGlyTrpAlaAlaValCys 254
      6810 CAGTTCCAGCGCGCGGAGCTCGTCCG----- 6784
      255 AspTyrGluArgTyrAlaHisAspMetGlyTyrAlaIleLeuThrAlaArgLysPro 273
      6783 -----GTCAAGAGATTCCGCTATCTGCTGCTGCTGCGCCAGCGCGCG 6742

```

```

/ PRIOR APPLICATION NUMBER: US 09/266,965
/ PRIOR FILING DATE: 1999-03-12
/ PRIOR APPLICATION NUMBER: US 08/624,447
/ PRIOR FILING DATE: 1996-08-19
/ PRIOR FILING DATE: 1994-10-06
/ PRIOR APPLICATION NUMBER: PCT/US94/11279
/ PRIOR FILING DATE: 1994-10-06
/ PRIOR APPLICATION NUMBER: US 08/133,963
/ PRIOR FILING DATE: 1993-10-07
/ NUMBER OF SEQ ID NOS: 145
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 76
/ LENGTH: 53500
/ TYPE: DNA
/ ORGANISM: Streptomyces lavendulae
/ US-10-267-255-76

Alignment Scores:
Pred. No.: 1e-32 Length: 53500
Score: 376.50 Matches: 101
Percent Similarity: 50.84% Conservative: 51
Best Local Similarity: 33.78% Mismatches: 94
Query Match: 26.07% Indels: 53
DB: 15 Gaps: 11

US-10-069-353A-8 (1-275) x US-10-267-255-76 (1-53500)
QY 2 LeuProGlyGlyAlaProThrSerGlnValAlaGlyGlnMetTyrAspLeuValThrPro 21
DB 7560 CTCGCCATGCGCCCTCACCCGCTCCGAGAGGCGGCGCTTCAGACCGGTTACCGCG 7501
QY 22 LeuLeuAspSerValAlaIleGlyProCysAlaIleHisIleGlyThrTrpGluAsn--- 40
DB 7500 CTGGAGCGCGCTCTCTCCGCGAG-----AACCTGCATCTTGCGGCTACTGGAGCTCCGCC 7447
QY 41 AspGlyArgAlaSerTrpGlnGlnAlaIleAspArgLeuThrAspLeuValAlaGluArg 60
DB 7446 GACAGCGAGTGCGCGCTGCGCGAGGCGACCGCGCTCACCAGCATGATGGCCGAGCGG 7387
QY 61 ThrValLeuAspGlyGlyValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80
DB 7386 CTGGCGATGCGCGCGCTCCGCGCTCCGAGCATTCGCGCTGCGCGCGCGCGCGCGCG 7327
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
DB 7326 GTACCGATCGCGCGCTCCGAGCGGCGCATGTCAAGGACATTCGGTAGCCATGAGCAG 7267
QY 101 ValAlaIleAlaIleAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
DB 7266 GTGCTCGCGGCGCAAGCGCTGCGCGAGGCGCGGCTGCCGACCGCGCGGCTTCCAG 7207
QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
DB 7206 CCGCGCGAGCGCATGAGCTCCCTTCGAGAGCGAGAGCTTCGAGCGCTCATCGCCCTC 7147
QY 141 GlnSerLeuLeuGlnMetSerGlnProAspArgAla-----IleArgGlnIleLeuArg 158
DB 7146 GAATGATATCATCAATG-----CCGAGCGCGCGCGAGCTCGCCAGGCTCGCGCG 7093
QY 159 ValLeuArgProGlyGlyIleLeuGlyValThrGluValAlaValAspArg----- 174
DB 7092 GTGCTCGCGCGCGCGCGCGCTGCTGCTGCTCAGCATTCCTTGAGGCGCGCGCGCG 7033
QY 175 ---GluAlaGlyGlyIleMetPro-----ValSer 183
DB 7032 CCGAGGCGCGCGCGCGCGCTCCAGCGCTACCTCCAGCATTCATGATGACCATGTGTACG 6973
QY 184 GlyAspArgTrpProThrGlyLeuArg-----IleCys 194
DB 6972 GCGGAGCGGTACCTCCCTCGCTGCGGCGGCGCGCGCTGCGTGGAGGAGTTCTCCGAC 6913
QY 195 LeuAlaGlnGlnLeuLeuGlnSerLeuArgAlaAlaGlyPheGlnIleLeuAspTrpGln 214
DB 6912 ATCAGCAGCACGACCTCGAGAAAGACC-----TTCAGGCTGCTTCG---GAG 6868

```



```
QY      6 AlaProthrSerGlnGlnValGlyGlnMetTyrAspLeuValThrProLeuAsnSer 25
      13 GGGTCGGCTCCGACCGGATGTCGAACTCTACACCGGTCACCGACTTGATGTCAC 72
QY      26 ValAlaGlyGlyProCysAlaIleHisGlyTyrTyrGlnAsn--AspGlyArgAla 44
      73 GCGGAAGCGGCGC-----TACATGCACGGGTGCTACTGGGCGGAGCCGACGTCCCGCAG 126
QY      45 SerTyrGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGlnArgThrValLeuAsp 64
      127 ACGGTGAAGAGGACGAGCGACCGCGTCGACCGACTACGTCTCGAGCGCTCGCGCTCGCC 186
QY      65 GlyGlyValArgLeuAspValGlyCysGlyThrGlyGlnProAlaLeuArgValAla 84
      187 CCGCGGAGACGGGTGCTCGACGTGGGGTGGGCAAGGAGCCGCTTGGCGACTCGCC 246
QY      85 ArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGlnValAlaIleAla 104
      247 GCCCGGACGCGGGTGGCGGCCACCGGGGTCTCCATCAACCCCTACAGGAGGTGGTGTGCG 306
QY      105 AlaAspCysAlaArgGlnArgGlyLeuSerHisArgValAspPheSerCysValAspAla 124
      307 CCGGACGCTCGCGAGAGGAGGCGC--GACGAGGCGACCGAGTTCGCATCGGTACATG 363
QY      125 MetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTyrAlaMetGlnSerLeu 144
      364 CTCGGCTCCCTTCCGACCGGCTGCTCGACGCTGTAGCGCATCGAGACATCTGC 423
QY      145 GluMetSerGlnProAspArgAlaIleArgGlnIleLeuArgValLeuLysProGlyGly 164
      424 CACGCGCTCGAAAGCGGCGCACGCTCTTCAACGAGATCGCCGGGTGCTGGCGCGCGCG 483
QY      165 IleLeuGlyValThrGlnValValLysArgGlnAlaGlyGlyMetProValSer--- 183
      484 CCGGTGACGTGACGAGACTTCGTGCTGCGCGCG-----CCCTGAGCGGAC 528
QY      184 -----GlyAspArgTyrProThrGlyLeuArgIleCys 194
      529 GGGTCAGAGCATGCTGCAACCGGCAAGCAACTTCACAGAGGCG---CCCGTCTC 585
QY      195 LeuAlaGlnGlnLeuLeuGlnSerLeuArgAlaAlaGlyPheGlnIleLeuAspTyrGln 214
      586 ACCCGGAGCGGTACGAGGACTGCAATCGGTGGGTGGGTGGGTGGAGTTCCTC 645
QY      215 AspValSerSerArgThrArgTyrPheMetProGlnPheAlaGlnLeuAlaHis 234
      646 GACATCGGGGACGAGGTGCGG-----CCCTCTAC--GAGCGGTGGCGGAG 693
QY      235 GlnHisGlyIleAlaAspArgTyrGlyProAlaValAlaGlyTyrAlaAlaValCys 254
      694 ATGCGTGGCGGACGAGGAGCGGTGCGCTCCACATG----- 729
QY      255 AspTyrGlnLysTyr-----AlaHisAsp 262
      730 GACGACGAGCGGTTCACCGCATGTGTCAGCGGCATCGACCGCATGGGTGGTGAGAG 789
QY      263 MetGlyTyrAlaIleLeuThrAlaArgLysPro 273
      790 GTGGGTACTCGGTGTGTCACCGGCGGAAACCG 822
Db
RESULT 14
US-10-267-255-35
; Sequence 35, Application US/10267255
; Publication No. US20030124689A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
```

```

; CURRENT APPLICATION NUMBER: US/10/267,255
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 35
; LENGTH: 828
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-10-267-255-35

Alignment Scores:
Pred. No.: 2,836-31      Length: 828
Score: 345.00      Matches: 89
Percent Similarity: 49.83%      Conservative: 56
Best Local Similarity: 30.58%      Mismatches: 102
Query Match: 23.89%      Indels: 44
DB: 15      Gaps: 10

US-10-069-353A-8 (1-275) x US-10-267-255-35 (1-828)
QY      6 AlaProthrSerGlnGlnValGlyGlnMetTyrAspLeuValThrProLeuAsnSer 25
      13 GGGTCGGCTCCGACCGGATGTCGAACTCTACACCGGTCACCGACTTGATGTCAC 72
QY      26 ValAlaGlyGlyProCysAlaIleHisGlyTyrTyrGlnAsn--AspGlyArgAla 44
      73 GCGGAAGCGGCGC-----TACATGCACGGGTGCTACTGGGCGGAGCCGACGTCCCGCAG 126
QY      45 SerTyrGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGlnArgThrValLeuAsp 64
      127 ACGGTGAAGAGGACGAGCGACCGCGTCGACCGACTACGTCTCGAGCGCTCGCGCTCGCC 186
QY      65 GlyGlyValArgLeuAspValGlyCysGlyThrGlyGlnProAlaLeuArgValAla 84
      187 CCGCGGAGACGGGTGCTCGACGTGGGGTGGGCAAGGAGCCGCTTGGCGACTCGCC 423
QY      85 ArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGlnValAlaIleAla 104
      247 GCCCGGACGCGGGTGGCGGCCACCGGGGTCTCCATCAACCCCTACAGGTGGTGTGCG 306
QY      105 AlaAspCysAlaArgGlnArgGlyLeuSerHisArgValAspPheSerCysValAspAla 124
      307 CCGGACGCTCGCGAGAGGAGGCGC--GACGAGGCGACCGAGTTCGCGATCGGTACATG 363
QY      125 MetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTyrAlaMetGlnSerLeu 144
      364 CTCGGCTCCCTTCCGACCGGCTGCTCGACGCTGTAGCGCATCGAGACATCTGC 423
QY      145 GluMetSerGlnProAspArgAlaIleArgGlnIleLeuArgValLeuLysProGlyGly 164
      424 CACGCGCTCGAAAGCGGCGCACGCTCTTCAACGAGATCGCCGGGTGCTGGCGCGCGCG 483
QY      165 IleLeuGlyValThrGlnValValLysArgGlnAlaGlyGlyMetProValSer--- 183
      484 CCGGTGACGTGACGAGACTTCGTGCTGCGCGG-----CCCTGAGCGGAC 528
QY      184 -----GlyAspArgTyrProThrGlyLeuArgIleCys 194
      529 GGGTCAGAGCATGCTGCAACCGGCAAGCAACTTCACAGAGGCG---CCCGTCTC 585
QY      195 LeuAlaGlnGlnLeuLeuGlnSerLeuArgAlaAlaGlyPheGlnIleLeuAspTyrGln 214
      586 ACCCGGAGCGGTACGAGGACTGCAATCGGTGGGTGGGTGGGTGGAGTTCCTC 645
QY      215 AspValSerSerArgThrArgTyrPheMetProGlnPheAlaGlnIleLeuAlaHis 234
```

```

Db      646 GACATGGGGAGCAGCTCGG-----CCCTCTAC---GAGCGGCGGCGGGAAG 693
QY      235 GlnHieGlyIleAlaAspArgTyrglyProAlaValAlaGlyTrrAlaAlaValCys 254
Db      694 ATGCGGCGGCGGCGGAGCGCTCGGCTCCACATG-----729
QY      255 AepTyrglyLysTyrr-----AlaHisAsp 262
Db      730 GACGACGAGCGGCTTCCACCGCATGTGACGCGCATGACCGCATGGGCTCGTGAGAG 789
QY      263 MetGlyTrrAlaIleuThrAlaArgLysPro 273
Db      790 GTGCGCTACTGGGTGGTCCGCGCGGAAACCG 822

RESULT 15
US-10-496-377-1
; Sequence 1, Application US/10496377
; Publication No. US20050118590A1
; GENERAL INFORMATION:
; APPLICANT: Piel, Jörn
; TITLE OF INVENTION: Novel Gene Cluster of Pederin Biosynthesis Genes
; FILE REFERENCE: 033285-020
; CURRENT APPLICATION NUMBER: US/10/496,377
; PRIOR FILING DATE: 2004-05-24
; PRIOR APPLICATION NUMBER: PCT/EP02/13085
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: EP 01127395.0
; PRIOR FILING DATE: 2001-11-22
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 52659
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Bacterial
; OTHER INFORMATION: Symbiont of Paederus fuscipes
; FEATURE:
; OTHER INFORMATION: Sequence of the pederin biosynthetic gene cluster.
; OTHER INFORMATION: Differences in the sequence are dependent on
; OTHER INFORMATION: whether the DNA was sequenced from the cosmid or
; OTHER INFORMATION: the genome
; FEATURE:
; NAME/KEY: gene
; LOCATION: (274)..(1212)
; OTHER INFORMATION: Peda putative methyltransferase
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1226)..(2620)
; OTHER INFORMATION: Pedb putative FMN-dependent oxidoreductase
; FEATURE:
; NAME/KEY: gene
; LOCATION: (2786)..(3800)
; OTHER INFORMATION: Pedc putative acyltransferase
; FEATURE:
; NAME/KEY: gene
; LOCATION: (3919)..(5013)
; OTHER INFORMATION: Pedd putative acyltransferase
; FEATURE:
; NAME/KEY: gene
; LOCATION: (5243)..(6064)
; OTHER INFORMATION: Pede putative methyltransferase
; FEATURE:
; NAME/KEY: gene
; LOCATION: (6309)..(32114)
; OTHER INFORMATION: Pedf mixed type I polyketide synthase/nonribosomal
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6305)..(8489)
; OTHER INFORMATION: module 1 PKS
; FEATURE:

```

```

; NAME/KEY: misc feature
; LOCATION: (8586)..(11618)
; OTHER INFORMATION: module 2 NRPS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (11817)..(15236)
; OTHER INFORMATION: module 3 PKS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15381)..(20192)
; OTHER INFORMATION: module 4 PKS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (20334)..(25508)
; OTHER INFORMATION: module 5 PKS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (25686)..(29315)
; OTHER INFORMATION: module 6 PKS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29478)..(32114)
; OTHER INFORMATION: module 7 incomplete PKS
; FEATURE:
; NAME/KEY: gene
; LOCATION: (32142)..(33443)
; OTHER INFORMATION: Pedg putative flavin-binding monooxygenase
; FEATURE:
; NAME/KEY: gene
; LOCATION: (33467)..(52267)
; OTHER INFORMATION: Pedh mixed type I polyketide synthase/nonribosomal
; OTHER INFORMATION: peptide synthetase
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (33467)..(33863)
; OTHER INFORMATION: module 1 incomplete PKS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (34070)..(38558)
; OTHER INFORMATION: module 2 PKS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (38678)..(42209)
; OTHER INFORMATION: module 3 PKS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (42314)..(45836)
; OTHER INFORMATION: module 4 PKS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45944)..(48128)
; OTHER INFORMATION: module 5 PKS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (48336)..(52267)
; OTHER INFORMATION: module 6 NRPS
; US-10-496-377-1

Alignment Scores:
Pred. No.: 1,03e-19
Score: 268.00
Percent Similarity: 49.47%
Best Local Similarity: 28.83%
Query Match: 18.56%
DB: 21
Length: 52659
Matches: 81
Conservative: 58
Mismatch: 109
Indels: 33
Gaps: 10

US-10-069-353a-8 (1-275) x US-10-496-377-1 (1-52659)
QY      10 GlnGlnValGlyGlnMetTyrrAp-----LeuValThrProIeuLeuAsnSer 25
Db      5267 GAGAAAGTGGCTTACGTTATATGATTCGCGAAGGCGCAAGGCGCTTATTTA-----5320
QY      26 ValAlaGlyGlyProCysAlaIleHisHisGlyTyrrTrp---GluAsnAspGlyArgAla 44

```

```
Db      5321 ----TTGGGCGGA-----CATATGCAATGGGCGCTATTGGAGTGAAGTCACTGGGCGAGGG 5371
Qy      45 SerTrpGlnGluAlaIleAspArgLeuThrAspLeuValAlaGluArgThrValLeuAsp 64
Db      5372 AACCTTCCACAGCAGCAGAAAGGCTTGCGCAGATTATGATTGCTTAAGCACTTAAAG 5431
Qy      65 GtGtYValArgLeuLeuAspValGtYCySGtYThrGtYGlnProAlaLeuArgValAla 84
Db      5432 GCCGGGCATAAATTCTACGACATGGGATGTGGCTTGAGAGCTGCTGCTGAAGCTGGCC 5491
Qy      85 ArgAspAsnAlaIleGlnIleThrGtYIleThrValSerGlnValGlnValAlaIleAla 104
Db      5492 AAAGCCAAAGGTTGTTGTTGTTGATGGAATTAACCATTAAGCAAGAGCAACAGCTCAGTGCA 5551
Qy      105 AlaAspCyAlaArgGluArgGtYLeuSerHisArgValAspPheSerCysValAspAla 124
Db      5552 ATTACTCGGGCTGAGCAGACGCACTGCAAGAGCGGTTCGATTATTCATGAGTACGGCA 5611
Qy      125 MetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMetGlnSerLeuLeu 144
Db      5612 CTGAATATACCGTGGCAGACCAATCGTATGATGATGTTGTTGTTTTCGAGTCCATTTT 5671
Qy      145 GluMetSerGluProAspArgAlaIleArgGtYIleLeuArgValLeuLysProGtYGtY 164
Db      5672 CATATGGGTCTAT--AGAAAGGCTTTCATGAGGCTGCGCGAGTACTCAAGCCTGGTCT 5728
Qy      165 IleLeuGtYValThrGtYValValLysArgGluAlaGtYGtYLeuMetProValSerGtY 184
Db      5729 ACTTATTTATTAACGATTGCTCTT-----CTTCTGAGAGTACC 5770
Qy      185 AspArg-----TrpPro-ThrGtYLeuArgIle-CysLeuAlaGlnIle 199
Db      5771 GAAGCTTTTAAGAGATTGCTTGGAACACATTTCGCTTGTGCTCTCGCGAAGAT 5830
Qy      199 euLeu-GluSerLeuArgAlaAlaGtYPheGtYIleLeuAspTrpGluAspValSerSer 218
Db      5831 TATCTGAGCTATTAGCCGAGCGGAGTTCGAACTCATCAATAATGATGATATTACTGAT 5890
Qy      219 ArgThrArg--TyrPheMetProGlnPheAlaGluGluLeuAlaAlaHisGlnHisGtY 237
Db      5891 AATGTCATGCTCCTGCGCTGAGGCCCAAGTTGAAGAGAGCGATCGAACTGCATGACCGCAG 5950
Qy      238 -----IleAlaAspArgTyrGtYProAlaValAlaGtYTrpAlaAlaAlaVal 253
Db      5951 GTCGAGGCGCATTAATCTCATATGATACGAGAAGCGCATCGACGATGGCTTATCTG-- 6007
Qy      254 CysAspTyrGtYLeuYserYTrpAlaHisAspMetGtYTrpAlaIleLeuThrAlaArgLys 272
Db      6008 -----TTGAGATGATGAGCGAAGAACTTAGGCTACATGATGCTGATGGCTAAAG 6058
```

Search completed: August 29, 2005, 23:46:40  
Job time : 5041 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 29, 2005, 21:22:55 ; Search time 4680 Seconds

(without alignments)  
2847.262 Million cell updates/sec

Title: US-10-069-353A-8

Perfect score: 275

Sequence: 1 VLBGAPTSQVQGMVDVLT.....YKXAHMKGXAITTARKPVG 275

Scoring table:  
OLIGO Xgapop 60.0 , Xgapext 60.0  
Xgapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9408497

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O/cg12.1/USPTO.spool/US10069353/runat.26082005.172151.15052/app.query.fasta\_1.455  
-DB=GenEmbl -QFMT=fastap -SUFFIX=xge -MIMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR.SCOR=quality -THR.MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10069353 @CGN.1.1.3731 @runat.26082005.172151.15052 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG.SCOR=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -MARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_cm.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	275	100.0	828	AX089422	AX089422 Sequence
2	275	100.0	45624	AX089419	AX089419 Sequence
3	275	100.0	50000	AX089416	AX089416 Sequence
4	275	100.0	80161	AY007564	AY007564 Saccharop

5	275	100.0	80161	6	AR165018	AR165018 Sequence
6	275	100.0	80161	6	BD137649	BD137649 Biosynthe
7	275	100.0	80161	6	AR281866	AR281866 Sequence
8	37	13.5	36538	6	AX600587	AX600587 Sequence
9	10	3.6	738	6	AR534398	AR534398 Sequence
10	10	3.6	738	6	AX573660	AX573660 Sequence
11	10	3.6	788	8	BT009270	BT009270 Triflicum
12	10	3.6	1394	5	AY491379	AY491379 Acanthopa
13	10	3.6	6085	6	AE0304	AE0304 Sequence 4
14	10	3.6	6085	6	AR144763	AR144763 Sequence
15	10	3.6	10802	1	AE005764	AE005764 Caulobact
16	10	3.6	22399	1	AB088119	AB088119 Streptomy
17	10	3.6	90348	1	AF497482	AF497482 Micromono
18	10	3.6	297850	1	AP006577	AP006577 Gloeobact
19	10	3.6	298550	1	AP005961	AP005961 Bradyrhiz
20	10	3.6	298750	1	AP005375	AP005375 Thermosyn
21	9	3.3	400	6	AR356861	AR356861 Sequence
22	9	3.3	400	6	AR358417	AR358417 Sequence
23	9	3.3	759	6	AK619480	AK619480 Sequence
24	9	3.3	2191	5	AB034983	AB034983 Xenopus 1
25	9	3.3	10993	1	AE011153	AE011153 Methanosa
26	9	3.3	11466	6	AR354326	AR354326 Sequence
27	9	3.3	11466	6	AR355882	AR355882 Sequence
28	9	3.3	50049	1	AP003358	AP003358 Contination (4 of
29	9	3.3	54587	2	AC146908	AC146908 Contination (4 of
30	9	3.3	64492	1	AB086653	AB086653 Streptomy
31	9	3.3	66201	1	AP003601	AP003601 Nostoc sp
32	9	3.3	95684	2	AC134115	AC134115 Rattus no
33	9	3.3	98835	9	AC073195	AC073195 Homo sapi
34	9	3.3	110000	1	AB017180	AB017180 Contination (17 of
35	9	3.3	110000	1	BX571856	BX571856 Contination (3 of
36	9	3.3	110000	1	BX571856	BX571856 Contination (3 of
37	9	3.3	110000	1	BX571857	BX571857 Contination (3 of
38	9	3.3	132981	8	AC147802	AC147802 Oryza bat
39	9	3.3	152950	2	AC121067	AC121067 Rattus no
40	9	3.3	160031	4	AC147681	AC147681 Cantis Fam
41	9	3.3	164398	10	AC116808	AC116808 Mus muscu
42	9	3.3	169065	2	AP002338	AP002338 Homo sapi
43	9	3.3	201472	2	AC073924	AC073924 Homo sapi
44	9	3.3	216120	9	AC098973	AC098973 Homo sapi
45	9	3.3	223337	10	AL928564	AL928564 Mouse DNA

#### ALIGNMENTS

RESULT 1  
AX089422 828 bp DNA linear PAT 21-MAR-2001  
LOCUS Sequence 7 from Patent WO0116303.  
DEFINITION AX089422  
ACCESSION AX089422.1 GI:13443683  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Saccharopolyspora spinosa  
Saccharopolyspora spinosa  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.  
REFERENCE  
AUTHORS Ebers, G., Moehle, V., Froede, R., Velten, R. and Salas, V.A.  
TITLES Nucleic acids which code for the enzyme activities of the spinosyn  
biosynthesis  
JOURNAL Patent: WO 0116303-A 7 08-MAR-2001;  
BAYER AG (DE)  
FEATURES  
source  
1..828 Location/Qualifiers  
/organism="Saccharopolyspora spinosa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:60894"  
1..828  
/note="unnamed protein product; ORF1, O-Methyltransferase"

CDS  
1..828  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC34874.1"  
/db\_xref="GI:13443684"

## ORIGIN

/translation="MLPGAPTSQGVQWMDVLTPLNSVAGGPCAIHHGWENDGRA  
SMOQADRLTDLVDAERTVLDGVRVLAIDGGGTGCPALPAPADMAIOTGTIVSOYVA  
IADQABREGLSHRVDRESCVDAMSILPYPNAPAPDAAMSLMSPDAIETILVTL  
KRGILGIVTBVYRKAGGGMPSVGBDMFPGRLCLAEQLLESIRAAQFEILMEDVSS  
RRTYFMPQAEELAHQHGLADRYGPVAVAGMAAADCYERKATIDWGYALTTPKPVG"

## Alignment Scores:

Pred. No.:	4,63e-292	Length:	828
Score:	275.00	Matches:	275
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-069-353a-8 (1-275) x AX089422 (1-828)

```
QY 1 ValLeuProGlyValAlaProThSerGlnGlnValGlyGlnMetTyAspLeuValThr 20
Db 1 GTGTTCAGTGGCCGACCAACATCGACAGCTGGGAGATGATGACCTGTCAAG 60
QY 21 ProLeuLeuAsnSerValAlaGlyProCysAlaIleHisGlyTyTTPGluAsn 40
Db 61 CGGTCTGTAACCTGGTCGGGGGGCCCTCGCCATCCACAGCGCTACCTGGGAGAAC 120
QY 41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
Db 121 GACGGCGGGCTTCCTGGCAGCAGCGCCGACCGCTCAACGACTTGTGCGCGAAGC 180
QY 61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80
Db 181 ACCGTCTCGATGGCGCGCTTCGACTGCTGATGGGGGTGGCGGTACCGGACCAACGACG 240
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyTleThrValSerGlnValGln 100
Db 241 CTGGCGGTGGCGGCGACACCGGATCCAGATCACCGGATCCCTCGACCGCAGGTGCA 300
QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
Db 301 GTGGCCATGGCGCTGATTTGGCCAGCAACCGGACCTAAGCAGCGGGGTGACCTTCTG 360
QY 121 CysValAspAlaMetSerLeuProTyProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
Db 361 TGGTGATGATGCCATGTCCTCGCGTACCGGACAAATGCTTTCGACCGCTGGGGCATG 420
QY 141 GlnSerLeuLeuGlnMetSerGluProAspArgAlaIleArgGlnIleLeuArgValLeu 160
Db 421 CAGTCCCTGTTGGAGATGTCGCAACCGGACCTGCAATCCGGAAATCTTGCAGTACTC 480
QY 161 LysProGlyGlyTleLeuGlyValThrGluValAlaIleArgGluAlaGlyGlyMet 180
Db 481 AAACCCGGTGGCATCTCCGGCGTCACGAGGTGTCAACGAGAACGGGGCGGGGAGATG 540
QY 181 ProValSerGlyAspArgTrpProThrGlyLeuArgTleCysLeuAlaGlnLeuLeu 200
Db 541 CCGGTGTCGGGGACGATGGCCGACCGGCTTCGATCTGCTGCTGCTGACCAACTTTCG 600
QY 201 GluSerLeuArgAlaIleGlyPheGlnIleLeuAspTrpGluAspValSerSerArgThr 220
Db 601 GAATCCCTGCTGGACGGGGGTTGAGATCTCGATTTGGAGAGCACTGTGCTGAGAGACC 660
QY 221 ArgTyTyrPheMetProGlnPheAlaGluGluLeuAlaAlaHisGlnHisGlyTleAlaAsp 240
Db 661 CGGTACTTTCATGCGCGAGTTCGCCGAAGAGCTCGCTGGCCGACACGAGGAGTCGGGAC 720
QY 241 ArgTyTyrGlyProAlaValAlaGlyTyrPalAlaAlaValCysAspTyTyrGluTyTyrAla 260
Db 721 AGGTACGGGCGCGCTGCTGCGCGCTGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
QY 261 HisAspMetGlyTyTyrAlaIleLeuThrAlaArgLysProValGly 275
Db 781 CACGACATGGGCTATGCGATTCTGACGGCGCGGAGACCGGCTCGGC 825
```

## RESULT 2

AX089419/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

LOCATION/Qualifiers

/organism="Saccharopolyspora spinosa"

/mol\_type="unassigned DNA"

/db\_xref="taxon:60894"

ORIGIN

Alignment Scores:

Pred. No.:	1.73e-290	Length:	45624
Score:	275.00	Matches:	275
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-069-353a-8 (1-275) x AX089419 (1-45624)

```
QY 1 ValLeuProGlyValAlaProThSerGlnGlnValGlyGlnMetTyAspLeuValThr 20
Db 828 GTGTTCAGTGGCCGACCAACATCGACAGCTGGGAGATGATGACCTGTCAAG 769
QY 21 ProLeuLeuAsnSerValAlaGlyProCysAlaIleHisGlyTyTTPGluAsn 40
Db 768 CGGTCTGTAACCTGGTCGGGGGGCCCTCGCCATCCACAGCGCTACCTGGGAGAAC 709
QY 41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
Db 708 GACGGCGGGCTTCCTGGCAGCAGCGCCGACCGCTCAACGACTTGTGCGCGAAGC 649
QY 61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80
Db 648 ACCGTCTCGATGGCGCGCTTCGACTGCTGATGGGGGTGGGTACCGGACCAACGACG 589
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyTleThrValSerGlnValGln 100
Db 588 CTGGCGGTGGCGGCGACACCGGATCCAGATCACCGGACCTGCAACCGGAGGTGCA 529
QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
Db 528 GTGGCCATGGCGCTGATTTGGCCAGCAACCGGACCTAAGCAGCGGGGTGACCTTCTG 469
QY 121 CysValAspAlaMetSerLeuProTyProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
Db 468 TGGTGATGATGCCATGTCCTCGCGTACCGGACAAATGCTTTCGACCGCTGGGGCATG 409
QY 141 GlnSerLeuLeuGlnMetSerGluProAspArgAlaIleArgGlnIleLeuArgValLeu 160
Db 408 CAGTCCCTGTTGGAGATGTCGCAACCGGACCTGCAATCCGGAAATCTTGCAGTACTC 349
QY 161 LysProGlyGlyTleLeuGlyValThrGluValAlaIleArgGluAlaGlyGlyMet 180
Db 348 AAACCCGGTGGCATCTCCGGCGTCACGAGGTGTCAACGAGAACGGGGCGGGGAGATG 289
QY 181 ProValSerGlyAspArgTrpProThrGlyLeuArgTleCysLeuAlaGlnLeuLeu 200
Db 288 CCGGTGTCGGGGACGATGGCCGACCGGCTTCGATCTGCTGCTGCTGCTGCTGCTGCTGCTG 229
```



QY 201 GluSerLeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220  
 Db 228 GAATCGCTGCGTGAAGCGGGGTTTCGAGATCTCGATTGGAGAGACGGTGTGCGAGAGACC 169  
 QY 221 ArgTyrPheMetProGlnPheAlaGluGluLeuAlaAlaGlnHisGlnHisGlyIleAlaAsp 240  
 Db 168 CGGTACTTCAATGCCGAGTTCGCGGAGAGAGCTGCTGCGACACAGACGCGGATCGCGGAC 109  
 QY 241 ArgTyrGlyProAlaValAlaGlyTyrPAlaAlaAlaValCysAspTyrGluValTyrAla 260  
 Db 108 AGGTACGGGCGCGGTGTCGCGGCTGGCGCGCGGCTGCGGATTATGAGAAATATATGCC 49  
 QY 261 HisAspMetGlyTyrAlaAlaIleLeuThrAlaArgLysProValGly 275  
 Db 48 CACGACATGGGCTATGCGATTCTGACGCGCGGAGCGGCGGTGCGC 4  
 RESULT 3  
 LOCUS AX089416 50000 bp DNA linear PAT 21-MAR-2001  
 DEFINITION Sequence 1 from Patent WO0116303.  
 ACCESSION AX089416  
 VERSION AX089416.1 GI:13443677  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Saccharopolyspora spinosa  
 Saccharopolyspora spinosa  
 Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;  
 Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.  
 REFERENCE  
 AUTHORS Eberz, G., Moehrle, V., Froede, R., Velten, R. and Salaas, J. A.  
 TITLE Nucleic acids which code for the enzyme activities of the spinosyn  
 biosynthesis  
 JOURNAL Patent: WO 0116303-A 1 08-MAR-2001;  
 BAYER AG (DE)  
 FEATURES  
 source 1. 50000  
 location/Qualifiers  
 /organism="Saccharopolyspora spinosa"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:60894"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1 886-290 Length: 50000  
 Score: 275.00 Matches: 275  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0  
 US-10-069-353a-8 (1-275) x AX089416 (1-50000)  
 QY 1 ValLeuProGlyAlaAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20  
 Db 44797 GTGTGGCAGGTGGCGGACCAACATCGCAGAGTTGGCAGATGATGACCTGTCACG 44856  
 QY 21 ProLeuLeuAsnSerValAlaGlyGlyProCysAlaIleHisHisGlyTyrTrpGluAsn 40  
 Db 44857 CCGTTGCTGAACCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 44916  
 QY 41 AspGlyValArgLysSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60  
 Db 44917 GACGGGCGGCGCTTCTTGCGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 44976  
 QY 61 ThrValLeuAspGlyValAlaGlyLeuLeuAspValGlyCysGlyTyrTrpGluProAla 80  
 Db 44977 ACCGTGCTCATGGCGGCGGCTTCGACCTGCTGATGTGGGTGCGGTACCGGACCAACGCG 45036  
 QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100  
 Db 45037 CTGGCGCTTCGCGCGGCGGCAACGCGATCCAGATCCCGGCAATCCGTCAGCCGAGGTGCA 45096  
 QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120

Db 45097 GTGGCCATGCGCGCTGATTGGCGGACGCGGAGCGGACCTAAGCCACCGGCTGACCTTCG 45156  
 QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140  
 Db 45157 TCGGTGAGGATGATGCTCCCTGCGGATCCCGGCAATGCTTTCGAGCGCGCGCGGCG 45216  
 QY 141 GlnSerLeuLeuGlnMetSerGlyProAspArgAlaAlaArgGlnIleLeuArgValLeu 160  
 Db 45217 CAGTCGCTTGGAGATGTCGCAACCGGACCGGTCGATCCGGAATTCCTTGAGATATC 45276  
 QY 161 LysProGlyGlyIleLeuGlyValThrGluValAlaLysArgGluAlaGlyGlyMet 180  
 Db 45277 AAACCGGTGGCATCTCGCGGCTCAACCGAGTGTGCAAAAGAGAGCGGCGCGGCGGAG 45336  
 QY 181 ProValSerGlyAspArgTrpProThrGlyLeuAlaGlyIleCysLeuAlaGluIleLeu 200  
 Db 45337 CCGGTGTCGGGAGACGAGTGGCGGACCGGCTTCGATCTGCTGATGAGCAACTTCG 45396  
 QY 201 GluSerLeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220  
 Db 45397 GAATCGCTGCGTGAAGCGGGGTTTCGAGATCTCGATTGGAGAGACGTTGTGCGAGACC 45456  
 QY 221 ArgTyrPheMetProGlnPheAlaGluGluLeuAlaAlaGlnHisGlnHisGlyIleAlaAsp 240  
 Db 45457 CGGTACTTCAATGCCGAGATTCGCGGAGAGCTCGCTGCGCACACGACGCGGATCGCGGAC 45516  
 QY 241 ArgTyrGlyProAlaValAlaGlyTyrPAlaAlaAlaValCysAspTyrGluValTyrAla 260  
 Db 45517 AGGTACGGGCGCGGTGTCGCGGCTGGCGCGGCGGCTGCGGATTATGAGAAATATATGCC 45576  
 QY 261 HisAspMetGlyTyrAlaAlaIleLeuThrAlaArgLysProValGly 275  
 Db 45577 CACGACATGGGCTATGCGATTCTGACGCGCGGAGCGGACCGGTGCGC 45621  
 RESULT 4  
 LOCUS AY007564 80161 bp DNA linear BCT 16-MAY-2001  
 DEFINITION Saccharopolyspora spinosa probable lysR-type transcriptional  
 regulator, probable keto acyl reductase, spinosad biosynthetic gene  
 cluster, complete sequence, and probable exodeoxyribonuclease V  
 genes, complete cds; and unknown gene.  
 ACCESSION AY007564 GI:13162633  
 VERSION AY007564  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Saccharopolyspora spinosa  
 Saccharopolyspora spinosa  
 Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;  
 Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.  
 REFERENCE  
 AUTHORS Waldron, C., Matsushina, P., Rostek, P. R. Jr., Broughton, M. C.,  
 Turner, J., Madduri, K., Crawford, K. P., Merlo, D. J. and Balz, R. H.  
 TITLE Cloning and analysis of the spinosad biosynthetic gene cluster of  
 Saccharopolyspora spinosa  
 JOURNAL Chem. Biol. 8 (5), 487-499 (2001)  
 MEDLINE 21257765  
 PUBMED 11358695  
 REFERENCE  
 AUTHORS 2 (bases 1 to 80161)  
 Turner, J., Madduri, K., Crawford, K. P., Merlo, D. J. and Balz, R. H.  
 TITLE Direct Substitution  
 JOURNAL Submitted (18-AUG-2000) Dow Agrosciences, 9330 Zionsville Rd.,  
 Indianapolis, IN 46268, USA  
 FEATURES  
 source 1. 80161  
 location/Qualifiers  
 /organism="Saccharopolyspora spinosa"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:60894"  
 complement(1135..1971)  
 CDS  
 /note="ORF L16"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable lysR-type transcriptional regulator"  
 /protein\_id="AAG23282.1"

gene  
CDS  
/db\_xref="GI:13162654"  
/translation="METRELRYPVAVAEELHFGRAQRLLAQPLSRTIAOLERG  
VILORTSRKVSITTEAGAMLTREGRAIILGALAAEGRATSOPIVLAAGASG  
ELLAKLLDAVAPGAVAVDILLCEQOPKTHIDGADYALILHQFPDPAELDIETLN  
TEQOVAIILPTSHPLASEPHVMADVSLDPLARWPGDYVPGGVEVNRQTLF  
QMIALGRTIVVMESSRVNMLEGLAVPVLDADVTYVIAWPHSRSRALAGLVVAT  
LI"  
2024. .2791  
/note="ORF L15"  
/codon\_start=1  
/transl\_table=11  
/product="probable keto acyl reductase"  
/protein\_id="AAG23281.1"  
/db\_xref="GI:13162653"  
/translation="WLVPRPRPATLESMEOTIALVTGANKGIGYEIAAGCALGMS  
VIGARDHORGEBAVAKLRADGVDAVASLDVTDASVAAAALLSEERAGRLDIVLNN  
AGIAGWPEPEPSTVTPASIRAYETNVISVYVTRNAMLPLIKSEPRRIYVNSSHYAS  
LTIQITPGVDLGGISGAYSPSKTFILAITTIQAKELSDTNIKINNACPGYVATDLNGF  
HGSTRPADGARIALRLATLPPDGPFGMFDGANVW"  
complement (3416. .4165)  
/gene="spn5"  
complement (3416. .4165)  
/note="probably involved in forosamine biosynthesis"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="probable N-dimethyltransferase"  
/protein\_id="AAG23280.1"  
/db\_xref="GI:13162652"  
/translation="MSRVSDTPEITSSVYSPPHADIYDAIHSARGDMAEAEEVVO  
VTRRLPEAGSLDVAACGTGAHLERFPAEYAKVAGLEISDAMEKRIARPEVPIHGD  
IEDFDGEEFDVITCLCTPTAAVWRVTDLRRVTRNARHLAAGVAVIEPMWPKFI  
DGFVDAVVAHGERVSRISVSVLEGRTSRMVTRVYVAPPTGIRDPTEIELSLFTE  
EXTALEDAGIAEVLPGANGRLFVGIRN"  
complement (4168. .5325)  
/gene="spn4"  
complement (4168. .5325)  
/note="probably involved in forosamine biosynthesis"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="probable aminotransferase"  
/protein\_id="AAG23279.1"  
/db\_xref="GI:13162651"  
/translation="MINIHOPIITGTEELDAIAEVPASNWIIGPPTRTFEAREAHILG  
VBPQVYPLANSGRNAELTYQVLDLGGPQDVILPSISFYAANAATASSGARVYFCVD  
PRTLNTLDVBARAITPATKAVLLHITGSGPESVTRIAIDFCRKGIMLTEDSACAVAS  
SVHGTACGTFFGDLATWSFDAMKILVTGDSGMFYAADPELAHRARRLAYAGLQMSGFD  
SAKSNRMDIRVEDIGORLIGNDVTALGSGVQLRKLPEFINRRRIAYALQYRLISDV  
PGVLLPPTLPDGVSSHVYVQLAPEIRDOVAQOMLERGITSYRYPPIHLVPIYRA  
DCKLPBAEDACRITLLPLPHSPDLDAEAVTVADDFQKAVHHIISGRSPLRK"  
complement (5363. .6751)  
/gene="spnQ"  
complement (5363. .6751)  
/note="probably involved in forosamine biosynthesis"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="probable NDP-hexose-3,4-dehydratase"  
/protein\_id="AAG23278.1"  
/db\_xref="GI:13162650"  
/translation="MOSKRTALGKGRARVTCDDTCATATEMVPDAKDRILASVRY  
HREBSPTFVAGSTPIRPSGAVLDEDDRYALVEALELPIAAGNARPRESPAPFG  
LKAHLVNGSSANLLASLSITSPKIGERKLPGCEVITAAVGPPTTINPAONKIVP  
VAVDELGTYNATPDRIKAAVETKTRAILMLATLIPFADSIABIAKHEHLFVEDN  
CVAVSGTYGRLLGTFGDLTVSPYAHNITSGEGGCYVLTGIELARIIESLIDWGRD  
CWCEPVDVTCRRKEDYHGLTLPYGDHKTYSHVGNLKTDLQALALISQLSKISA  
FSASRRMRRLREGSLGPGLLPVATPHSDPSMWGFAITTI SADAGFTRALVNPLE  
SPNIGTRLLFGNITRHPAFEDQVRYALDALNSDLVTDRTFWGVYREITDQMIDYV  
VSIABFVAKSS"

gene  
CDS  
7083. .8450  
/gene="spn4"  
7083. .8450  
/gene="spn"  
/note="Involved in forosamine addition"  
/codon\_start=1  
/transl\_table=11  
/product="probable NDP-forosamyItransferase"  
/protein\_id="AAG23277.1"  
/db\_xref="GI:13162649"  
/translation="WILGMLPGCSIAIEFNKRVLFTPLPASHFENLVPLAMALBAAG  
HEVRVAILCPNVASVNTGAGLTAVPVDELDIISAAKVELVSGSVSEDEGRHELF  
DELSINGRDTDAVEQLHLVDVDSLDMGFAEKMQPDLVVMAMCSGVVRAALG  
ARHVMVLAALDPSGWLRSGLFLEYGESKPEQGVDPGLWIAKAKAFGATDEELVTG  
QATIDPISMMRLPVLDLIDYIMRFPVYVGPVLPFWLBEREPTKRVCTTGLTKRRLS  
RVTREYGSOSIOEOAMVERLLRGARLDVEYIATISDQENEMELPSNVVHYEYPL  
NELBSCSVIITHGSTTTOETATYNGVQLILPGLTFMDESRAEILADRGGLVDDPA  
TFTEEDVGQARLLIDEPSFANAAALLIREIIESSPSPHDIVPRLEKVAERENRTGQ  
SDGHP"  
8967. .10427  
/gene="spnQ"  
8967. .10427  
/gene="spnQ"  
complement (11530. .12492)  
/note="involved in forosamine biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="probable NDP-hexose-2,3-dehydratase"  
/protein\_id="AAG23276.1"  
/db\_xref="GI:13162648"  
/translation="WSSSEVABASNAAPLGSNNTRRPVDSLALSCNKMIPTEHICWL  
ADRIEENSFETNRIIPFRLSKWKCDASTENI.VHADGPFYVEGLQVETNYGAAPSMMQ  
PIINQAEVIGIITVKEIDGVILHCLMSAKMPEGVNNVLTSLPQATSNYQAHRS  
VPPVADYELGAGRLVDVLOSQESGFYKRRNMVVEQVEVPLVPCMLTGLG  
VVALRQONINVMOTRTVLSCLPHDSATGPBELAASERPQAVASLSHGIDSSIS  
EAVGPEFEAKARVRLRATRVPLSVVDKRYRPTDTIAHODCKYPAIVASVATREVA  
SNTQPMITPEKQGEIATLVKRIIGVYHGLFARVBAQYKMAETAPTYQSVANYOST  
PSNDMPPLDVLVLTADPETVYESILSEEGRFYQAQRVRIIEVHEDFAARPSDFR  
MWTLGQLGELRLSTHFINIQARSIVASISHSLMALGR"  
10436. .11434  
/gene="spn4"  
10436. .11434  
/gene="spn"  
/note="Involved in forosamine biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="probable NDP-hexose-3-ketoreductase"  
/protein\_id="AAG23275.1"  
/db\_xref="GI:13162647"  
/translation="WRKPVRLGVLCGASFANRRMLPAMCDVAETEVAVASRDPAAKE  
RFAAFCEAVLGYORLLERPDIDAVVPLPGMNAEWIIGALBAKHVLAKEPLTTP  
ASDVARFELGLARRKRLRLRENYLPLHGRHDVVNDLQSGSIGELREFTAVFGIPLP  
DTRIRYRELGGALDDGVYPARAAHFLGLPVLVAGSIAESHOESDGLSGVDELQ  
SEGGVARIIDYGFVHNHRSAYEILGNSRRIYVDAFPPRAWQVIRIERGVVDELQ  
LPABDQVKATYARADIRACTGYDDRVAVDGSSESMIQQALYBALQARCCST"  
complement (11530. .12492)  
/gene="spn4"  
complement (11530. .12492)  
/note="involved in spinosyn aglycone biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="spn4"  
/db\_xref="GI:13162646"  
/translation="MPNVAAGSTVLPNIPMPREDRPIITFAVTHGLGSOVAPSYLIR  
TGTRETELIALVADRGVAIVITYTEGIGTCTTYTVGRQAGMADAAGAORPG  
SGLTITDCVGVGWTAQGQASAPFGEHLPTIAPBLRIAAAGAVPIDLLITHNDG  
VFTGVLDVAGLAAAYEPDLFPEBLITEAGTAVDQVEALAPLAVFLRLERELSDFL  
DVSGLFEQPRRARILAESVAGRNQGPVPTLVYSTDEIYPPAFGERLDSYBAAGT  
FVRMHPFSGMLHAPFAALASRVVSWPFDEHFSBSAISGPRDAR"  
complement (12696. .13547)  
/gene="spn1"  
complement (12696. .13547)

```

/gene="spnL"
/note="involved in spinosyn aglycone biosynthesis;
possibly involved in polyketide bridging"
/codon_start=1
/translation="methylesterase-like protein"
/product="methylesterase-like protein"
/protein_id="A03273.1"
/db_xref="GI:13162645"
/translation="MESIFDALHGRPLHHGTYAGCYREDAGTTPSDAODTDLFI
DKALRPGAHLPDGGGNGQPVVRAACASVAVTGIVNAQHLAAATRLANETGLGS
LEFDLVDQOLPYDPGFQAMAMQSVQVLDQAAIIRVHRIIBGGRVLDIITR
VRLPEEYAAVMTSTATLINSFALVSEAFELIEVTDLTQTRCMWSMYDELRLK
DELAGEPAAVGTYYQORYLGDIAAKGPPAQOLIAVAEYRKHPDYARNESGMFL
QARKOS"
complement(13592..14785)
/gene="spnK"
complement(13592..14785)

gene
CDS
complement(13592..14785)

Alignment Scores:
Pred. No.: 2,88e-290 Length: 80161
Score: 275.00 Matches: 275
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-069-353a-8 (1-275) x AY007564 (1-80161)

Qy 1 ValLeuProGlyValAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20
Db 20168 GTGTTGCGAGTGGCGGCGACCAATCGCAGAGTTGGGAGATGTATGACTGTGTCAGC 20227

Qy 21 ProLeuLeuAsnSerValAlaGlyGlyProCysAlaIleHisGlyTyrTrpGluAsn 40
Db 20228 CCGTTGCTGAATCTCGGTGCGGGGCGCCCTGCGCCATCCACACCGGCTACTGGAGAAC 20287

Qy 41 AspGlyValArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
Db 20288 GACGGGGCGGCTTCCTGCGAGACGCGGCGACCGGCTCACCGACTTGTGCGCGAACGG 20347

Qy 61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCysGlyTyrGlnProAla 80
Db 20348 ACCGTGCTGATGGCGGCGGCTTCGACTGCTGAGTGGGGTGGCGTACCGGACCAACGCG 20407

Qy 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
Db 20408 CTCGCGCTCGCGCGGCGACACCGGATCCAGATCACCGGCTACCGCTCAGCGGTGCA 20467

Qy 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
Db 20468 GTGGCCATCGCGCTGATGCGACGCGAACGCGGACTAAGCCACCGGGTGAATCTTCG 20527

Qy 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
Db 20528 TGGGTGATGATCCATGTCCTCGCGGTACCGGACATCTTTCAGCGCGCTGGGCATG 20587

Qy 141 GlnSerLeuGluMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160
Db 20588 CAGTCCGTGTGGAGATGTCGGAACCGGACCGGCTACCTTTCAGATCACTC 20647

Qy 161 LysProGlyValGlyIleLeuGlyValThrGluValAlaValAspArgGluAlaGlyValMet 180
Db 20648 AAACCCGGTGGCATCTCGCGGTGACCGAGGTGCTCAACGAGAGGGGCGGGGAGATG 20707

Qy 181 ProValSerGlyAspArgTrpProThrGlyLeuArgIleCysLeuAlaGluGlnLeuLeu 200
Db 20708 CCGGTGTCGGGGGACAGGTGGCGGACCGGCTTCGATCGCTGCTGAGCAACTTCG 20767

Qy 201 GluSerLeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220
Db 20768 GAATCGCTGCTGAGCGGGGTTCGAGATCTTCGATTTGGAGGACGCGTGTGAGAGACC 20827

Qy 221 ArgTyrPheMetProGlnPheAlaGluGluLeuAlaIleHisGlnHisGlyIleAlaAsp 240

```

```

Db 20828 CGGTACTTCATGCCGCGAGTTCGCCGAGAGGCTGCTGCGGACACAGACGGATCGCGGAC 20887

Qy 241 ArgTyrGlyProAlaValAlaGlyTyrAlaAlaAlaValCysAspTyrGluYsTyrAla 260
Db 20888 AGGTACGGGCGCGCTTCCTCCCGCTGCGGCGCGCGCTGCTGCGATATAGAAATATGCC 20947

Qy 261 HisAspMetGlyTyrAlaIleLeuThrAlaArgValProValGly 275
Db 20948 CACGACATGGGCTATCGATTCGACGCGCGGAGACCGGATCGGC 20992

RESULT 5
LOCUS AR165018 80161 bp DNA linear PART 17-OCT-2001
DEFINITION Sequence 1 from patent US 6274350.
ACCESSION AR165018
VERSION AR165018.1 GI:16238344
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 80161)
AUTHORS Baltz,R.H., Crawford,K.P., Broughton,M.Christine., Madduri,K.,
Merlo,D.J., Turner,J.R., Treadway,P.J. and Waldron,C.
TITILE Biosynthetic genes for spinosyn insecticide production
JOURNAL Patent: US 6274350-A 1 14-AUG-2001;
FEATURES
Location/Qualifiers
1..80161
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 2,88e-290 Length: 80161
Score: 275.00 Matches: 275
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-069-353a-8 (1-275) x AR165018 (1-80161)

Qy 1 ValLeuProGlyValAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20
Db 20168 GTGTTGCGAGTGGCGGCGACCAATCGCAGAGTTGGGAGATGTATGACTGTGTCAGC 20227

Qy 21 ProLeuLeuAsnSerValAlaGlyGlyProCysAlaIleHisGlyTyrTrpGluAsn 40
Db 20228 CCGTTGCTGAATCTCGGTGCGGGGCGCCCTGCGCCATCCACACCGGCTACTGGAGAAC 20287

Qy 41 AspGlyValArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
Db 20288 GACGGGGCGGCTTCCTGCGAGACGCGGCGACCGGCTCACCGACTTGTGCGCGAACGG 20347

Qy 61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCysGlyTyrGlnProAla 80
Db 20348 ACCGTGCTGATGGCGGCGGCTTCGACTGCTGAGTGGGGTGGCGTACCGGACCAACGCG 20407

Qy 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
Db 20408 CTCGCGCTCGCGCGGCGACACCGGATCCAGATCACCGGCTACCGCTCAGCGGTGCA 20467

Qy 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
Db 20468 GTGGCCATCGCGCTGATGCGACGCGAACGCGGACTAAGCCACCGGGTGAATCTTCG 20527

Qy 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
Db 20528 TGGGTGATGATCCATGTCCTCGCGGTACCGGACATCTTTCAGCGCGCTGGGCATG 20587

Qy 141 GlnSerLeuGluMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160
Db 20588 CAGTCCGTGTGGAGATGTCGGAACCGGACCGGCTACCTTTCAGATCACTC 20647

```

QY 161 LysProGlyGlyIleLeuGlyValThrGluValIleValysArgGluIleGlyGlyMet 180  
|||  
Db 20648 AAACCCGGAGGCACTCCCGCGTACCGAGGTGTGTAACGAGAACGGGGCGGAGAG 20707  
QY 181 ProValSerGlyAspArgTrpProThrGlyLeuArgIleCysLeuIleGluGlnLeu 200  
20708 CCGGTGTCCGGGAGACAGGTGTCCGACCGGCTTCGATCTGCTGGCTGAGCAACTTCG 20767  
QY 201 GluSerLeuArgAlaIleGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220  
20768 GAATCCTGCTGACACGGGGGTTCGAGATCTCGAATGGGAGGACGTGTCTGAGAGACC 20827  
QY 221 ArgTyrPheMetProGlnPheIleGluIleLeuAspTrpGluAspValSerSerArgThr 240  
20828 CGGTACTTCATGCCGCGAGTTCGCCGGAAGACTGTGCGCACGACGAGATCGCGGAC 20887  
QY 241 ArgTyrGlyProAlaValAlaGlyTyrAlaAlaAlaValCysAspTyrGluTyrAla 260  
20888 AGGTACGGGGCGGCTGTCCGCGGCTGGCGCGCGGTCTGGATTATGAGAAATATGCC 20947  
QY 261 HisAspMetGlyTyrAlaIleLeuThrAlaArgLysProValGly 275  
20948 CACGACATGGGCTATGCGATTTCGACGGCGCGGAGACCGGTCCG 20992  
RESULT 6  
BD137649 80161 bp DNA linear PAT 18-SEP-2002  
LOCUS BD137649 Biosynthetic genes for spinosyn insecticide production.  
DEFINITION BD137649  
ACCESSION BD137649.1 GI:23232594  
VERSION JP 2002505881-A/1.  
KEYWORDS Saccharopolyspora spinosa  
SOURCE Saccharopolyspora spinosa  
ORGANISM Saccharopolyspora spinosa  
REFERENCE Bacteria; Actinobacteria; Actinomycetales;  
Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.  
1 (bases 1 to 80161)  
AUTHORS Baltz,R.H., Broughton,M.C., Crawford,K.P., Madduri,K., Merlo,D.J.,  
Treadway,P.J., Turner,J.R. and Waldron,C.  
TITLE Biosynthetic genes for spinosyn insecticide production  
JOURNAL Patent: JP 2002505881-A 1 26-FEB-2002;  
DOW AGROSCIENCES LLC  
COMMENT OS Saccharopolyspora spinosa  
PN JP 2002505881-A/1  
PD 26-FEB-2002  
PE 16-FEB-1999 JP 2000535754  
PR 09-MAR-1998 US 09/036967  
PI RICHARD H BALTZ,R M CHRISTINE BROUGHTON,KATHERYN P CRAWFORD, PI  
KRISHNAMURTHY MADHURI,DONALD J MERLO,PATTI J TREADWAY,JAN R PI  
TURNER,  
PI CLIVE WALDRON  
PC C12N15/09,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P19/62,C12O1/  
PC 68//  
FC C07K14/195,C12N15/00,C12N5/00  
CC Biosynthetic genes for spinosyn insecticide production FH  
Key Location/Qualifiers  
FT source 1..80161  
FT /organism='Saccharopolyspora spinosa'.  
FEATURES  
source location/Qualifiers  
1..80161  
/organism="Saccharopolyspora spinosa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:60894"  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,88e-290 Length: 80161  
Score: 275.00 Matches: 275  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Db: 6 Gaps: 0

US-10-069-353a-8 (1-275) x BD137649 (1-80161)  
QY 1 ValLeuProGlyGlyAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20  
20168 GTGTGGCCAGGTGGGCGCACCAATCGACAGGTTGGGCAATGTATGACCTGTGCACG 20227  
QY 21 ProLeuLeuAsnSerValAlaGlyLysProCysAlaIleHisGlyTyrTrpGluAsn 40  
20228 CCGTCTCTAATCTCCGTCCGGGCGGCCCCCTCCGCATCCACACCGGCTACTGGAGAAC 20287  
QY 41 AspGlyArgAlaSerTrpGlnGlnAlaIleAspArgLeuThrAspLeuValAlaGluArg 60  
20288 GACGGGCGGGCTTCTTCGGCAGCAGGCCCGCGACCGGCTCACCGACTTGTCCGAAACG 20347  
QY 61 ThrValLeuAspGlyGlyValAlaArgLeuLeuAspValGlyCysGlyTyrGlyGlnProAla 80  
20348 ACCGTGCTCGATGGGCGGCTTGTGACTGTCTGATGTGGGGGTGGGATACCGGACACG 20407  
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100  
20408 CTGGCGGTCCGCGCGACACCGCATCCAGATCACCGGATCACCGTCAAGCCAGGTGCNA 20467  
Db 20468 GTGGCCATCGCGCTGATGCGCACGCGAAACCGGACTTAAGCACCGGGTGGACTTCTCG 20527  
QY 101 ValAlaIleAlaIleAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120  
20468 GTGGCCATCGCGCTGATGCGCACGCGAAACCGGACTTAAGCACCGGGTGGACTTCTCG 20527  
QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaIleTrpAlaMet 140  
20528 TCCGTGATGATGATGCTCTGCGTACCGGCAATGCTTTCGACGCGCGCTGGCCCATG 20587  
QY 141 GlnSerLeuLeuGlnMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160  
20588 CAGTGCTCTTGAGATGTCGAAACCGGACCGTGCATCCGGAAATCTTCGATCATCTC 20647  
QY 161 LysProGlyGlyIleLeuGlyValThrGluValValLysArgGluAlaGlyGlyMet 180  
20648 AAACCCGGGAGCATCTCGGCTTCACGAGGTCTGCAACGAGAACCGGCGCGGAGATG 20707  
QY 181 ProValSerGlyAspArgTrpProThrGlyLeuArgIleCysLeuIleGluGlnLeu 200  
20708 CCGGTGTCCGGGAGACAGGTGTCCGCGTTCGAGTCTGCTGAGCAACTTCG 20767  
QY 201 GluSerLeuArgAlaIleGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220  
20768 GAATGCTGCTGCGGACGCGGGGTTCGAGATCTTCGATGGAGAGACTGTGCGAGAAC 20827  
QY 221 ArgTyrPheMetProGlnPheAlaGluGlnLeuAlaIleHisGlyIleAlaAsp 240  
20828 CCGTACTTCATGCCGCGACAGTTCGCCGGAAGAGCTCGTGGCGCACACGCGGATCCG 20887  
QY 241 ArgTyrGlyProAlaValAlaGlyTyrAlaAlaAlaValCysAspTyrGluTyrAla 260  
20888 AGGTACGGGGCGGCTGTGCGCGGCTGGCGCGGCTGGATTATGAGAAATATGCC 20947  
QY 261 HisAspMetGlyTyrAlaIleLeuThrAlaArgLysProValGly 275  
20948 CACGACATGGGCTATGCGATTCTGACGGCGGGAACCGGTCCG 20992  
RESULT 7  
AR281866 80161 bp DNA linear PAT 10-APR-2003  
LOCUS AR281866 Sequence 1 from patent US 6521406.  
DEFINITION AR281866  
ACCESSION AR281866  
VERSION AR281866.1 GI:29717767  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 80161)  
AUTHORS Baltz,R.H., Broughton,M.C., Crawford,K.P., Madduri,K., Merlo,D.J.,  
Treadway,P.J., Turner,J.R. and Waldron,C.  
TITLE Spnd, a gene for spinosyn insecticide biosynthesis  
JOURNAL Patent: US 6521406-A 1 18-FEB-2003;

FEATURES  
source  
Location/Qualifiers  
1..80161  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Alignment Scores:  
Pred. No.: 2,886-290 Length: 80161  
Score: 275.00 Matches: 275  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-069-353a-8 (1-275) x AR281866 (1-80161)

QY 1 ValLeuProGlyValAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20  
DB 20168 GTGTTCAGAGTGGCGACCAACATCGACAGAGTTGGCAGATGATGACCTGTGCACG 20227  
QY 21 ProLeuLeuAsnSerValAlaGlyProCysAlaIleHisGlyTyrTrpGluAsn 40  
DB 20228 CCGTTCGTGAACCTCGGCGGGCGGCCCTGCGCCATCCACACGCGTACTGGAGAAC 20287  
QY 41 AspGlyValArgLaserTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60  
DB 20288 GACGGCGGGCTTCCTGCGACAGCGCCGCGGCTCACCACTTGTGCGCGAACGG 20347  
QY 61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCysGlyTyrTrpGluProAla 80  
DB 20348 ACCGTGCTCGATGGCGGCGTTGACATGCTGATGTGGGGTGGCGTACCGGACCAACGCG 20407  
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100  
DB 20408 CTCGGCGTCCGGCGCCACCAACGCGATCCAGATCACCGGCACTCCGACCGAGTGCAC 20467  
QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120  
DB 20468 GTGGCCATCGCCCTGATTCGCGACCGGAAACCGGACTAAGCCACCGGTTGGACTTCTCG 20527  
QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140  
DB 20528 TGGGTGATGTCATGTCCTCGCGGTACCGGACAACTCTTCACCGCGCGCTGGCGCATG 20587  
QY 141 GlnSerLeuLeuGlnMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160  
DB 20588 CAGTCCCTGTTGGAGATGTCGGAACCGGACCGTGCATCCGGGAAATCCTTCAGATGATC 20647  
QY 161 LysProGlyGlyIleLeuGlyValThrGluValLysArgGluAlaGlyGlyMet 180  
DB 20648 AAACCCGGTGGCATCTCGCGGTACCGAGGTGTCGAACGAGAAAGCGGCGGCGGAGATG 20707  
QY 181 ProValSerGlyAspArgTrpProThrGlyLeuArgIleCysLeuAlaGluGlnLeuLeu 200  
DB 20708 CCGGTGTCGGGGACAGGTGGCGGACCGGCTTCGATGTCGCTGGGTGAGCAACTTCTG 20767  
QY 201 GluSerLeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220  
DB 20768 GAATTCCTGCTGACAGCGGGGTTCGAGATCTCGATTTGGAGAGACGTTGCTGAGAGACC 20827  
QY 221 ArgTyrPheMetProGlnPheAlaGluGluLeuAlaAlaHisGlnIleGlyIleAlaAsp 240  
DB 20828 CCGTACTTCAATGCGCGAGTTGCGCGAAGACTCGCTGCGCACACGACCGGAGATCGCGGAC 20887  
QY 241 ArgTyrGlyProAlaValAlaGlyTyrAlaAlaAlaAlaValCysAspTyrGluIleTyrAla 260  
DB 20888 AGGTACGGGCGGCTGTCGCGCGGCTGGCGCGCGCTGCGCATTAATGAGAAATATGCC 20947  
QY 261 HisAspMetGlyTyrAlaIleLeuThrAlaArgLysProValGly 275  
DB 20948 CACGACATGGGCTATGAGATTGACGCGCGGAAAGCGGTCGCGC 20992

RESULT 8

AX600587/c  
LOCUS AX600587 36538 bp DNA linear PAT 14-FEB-2003  
DEFINITION Sequence 2 from Patent WO02079477.  
ACCESSION AX600587  
VERSION AX600587.1 GI:28400303  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Dow Agrosciences LLC (US)  
Patent: WO 02079477-A 2, 10-OCT-2002;  
Bio-synthetic genes for butenyl-spirosyn insecticide production  
Hahn, D. R., Jackson, J. D., Bullard, B. S., Gustafson, G. D., Waldron, C.  
and Mitchell, J. C.  
Pseudomonocardineae; Pseudomonocardineae; Saccharopolyspora.  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Saccharopolyspora sp. NRRL 30141  
Saccharopolyspora sp. NRRL 30141  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudomonocardineae; Pseudomonocardineae; Saccharopolyspora.

ORIGIN

Alignment Scores:  
Pred. No.: 7,336-29 Length: 36538  
Score: 37.00 Matches: 37  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 13.45% Indels: 0  
DB: 6 Gaps: 0

US-10-069-353a-8 (1-275) x AX600587 (1-36538)

QY 7 ProThrSerGlnGlnValAlaGlyGlnMetTyrAspLeuValThrProLeuLeuAsnSerVal 26  
DB 920 CCACATCCACAGAGTGGCGGACAGATGATGACCTGTGACCCCGTTCGATTCGATC 861  
QY 27 AlaGlyGlyProCysAlaIleHisGlyTyrTrpGluAsnAspGlyArg 43  
DB 860 GCGGGCGGCGCCCTGCGCCATCCACACGCGTACTGGGAGAACGCGGCGG 810

RESULT 9  
AR534398  
LOCUS AR534398 738 bp DNA linear PAT 08-OCT-2004  
DEFINITION Sequence 13 from patent US 6733998.  
ACCESSION AR534398  
VERSION AR534398.1 GI:53924595  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FATEURS  
source  
1..738  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Alignment Scores:  
Pred. No.: 1.06 Length: 738  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.64% Indels: 0  
DB: 6 Gaps: 0

US-10-069-353a-8 (1-275) x AR534398 (1-738)

Qy 68 ArgLeuLeuAspValGlyCysGlyThrGly 77  
|||||  
Db 121 CGGCTGCTCGACGTCGGCTGCCGACCGGC 150

RESULT 10  
AX573660  
LOCUS AX573660 738 bp DNA linear PAT 07-JAN-2003  
DEFINITION Sequence 13 from Patent WO02079465.  
ACCESSION AX573660  
VERSION AX573660.1 GI:27551330  
KEYWORDS  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1  
AUTHORS Thorson, J.  
TITLE Micromonospora echinospora genes encoding for biosynthesis of  
calicheamicin and self-resistance thereto  
JOURNAL Patent: WO 02079465-A 13 10-OCT-2002;  
MEMORIAL SLOAN-KETTERING CANCER CENTER (US)  
FEATURES  
source 1..738  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"  
/note="Bacteria"  
1..738  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD6016.1"  
/db\_xref="GI:27551331"  
/translation="MGPPEAAVEYENAYRGKSKWHDEADYADRIIARPDARLLD  
VCCGAGALETFATREPHVEGLAPAMALAHRLPGVRLAGDMRTFDLGVTFDAV  
TCLFTAVNLTGTVAEKRAVAAMASHLAAGVLTVEPPMFPRLFDGYGGLVREG  
RVAARVSRSTRGRTMRERLTNGDAIRFSQGLLTMTREBYDAAPAAAGCES  
ATVEGMLTRGILFVARTGHAHPITV"

CDS  
1..738  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD6016.1"  
/db\_xref="GI:27551331"  
/translation="MGPPEAAVEYENAYRGKSKWHDEADYADRIIARPDARLLD  
VCCGAGALETFATREPHVEGLAPAMALAHRLPGVRLAGDMRTFDLGVTFDAV  
TCLFTAVNLTGTVAEKRAVAAMASHLAAGVLTVEPPMFPRLFDGYGGLVREG  
RVAARVSRSTRGRTMRERLTNGDAIRFSQGLLTMTREBYDAAPAAAGCES  
ATVEGMLTRGILFVARTGHAHPITV"

Alignment Scores:  
Pred. No.: 1.06 Length: 738  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.64% Indels: 0  
DB: Gaps: 0

US-10-069-353a-8 (1-275) x AX573660 (1-738)

Qy 68 ArgLeuLeuAspValGlyCysGlyThrGly 77  
|||||  
Db 121 CGGCTGCTCGACGTCGGCTGCCGACCGGC 150

RESULT 11  
BT009270  
LOCUS BT009270 788 bp mRNA linear PLN 20-JUN-2003  
DEFINITION Trifolium aestivum clone wlx8.pk0008.g5:fls, full insert mRNA  
sequence.  
ACCESSION BT009270  
VERSION BT009270.1 GI:32128821  
KEYWORDS  
SOURCE Trifolium aestivum (bread wheat)  
ORGANISM Trifolium aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Trifolium.  
1 (bases 1 to 788)  
Tiney, S.V., Wolters, P., Powell, W., Dolan, M., Miao, G.-H.,  
Caraher, N.R., Hanafey, M.K. and Hanney, C.F.  
Direct Submission  
Submitted (20-JUN-2003) Crop Genetics, E. I. Dupont de Nemours and  
Company, 1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104,  
USA

FEATURES  
Location/Qualifiers

source 1..788  
/organism="Trifolium aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone="wlx8.pk0008.g5:fls"

ORIGIN  
Alignment Scores:  
Pred. No.: 1.12 Length: 788  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.64% Indels: 0  
DB: Gaps: 0

US-10-069-353a-8 (1-275) x BT009270 (1-788)

Qy 148 GIUProAspArgAlIleArgGluIleLeu 157  
|||||  
Db 541 GAGCCAGACCGTCGATTCGGAGATTGTG 570

RESULT 12  
AY491379/c  
LOCUS AY491379 1394 bp mRNA linear VRT 17-SEP-2004  
DEFINITION Acanthopagrus schlegelii FTZ-F1 (f1b) mRNA, partial cds.  
ACCESSION AY491379  
VERSION AY491379.1 GI:45686254  
KEYWORDS  
SOURCE Acanthopagrus schlegelii (black sea bream)  
ORGANISM Acanthopagrus schlegelii  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;  
Percoidae; Sparidae; Acanthopagrus.  
1 (bases 1 to 1394)  
Lin, X., Liang, B. and Zhang, S.  
Sequence and expression of cytochrome P450 aromatase and FTZ-F1  
genes in the procaridous black porgy (Acanthopagrus schlegelii)  
Gen. Comp. Endocrinol. 138 (3), 247-254 (2004)

JOURNAL  
PUBMED 15364207  
AUTHORS Xusheng, L., Bing, L. and Shuyi, Z.  
DIRECT SUBMISSION  
Submitted (28-NOV-2003) Institution of Zoology, Chinese Academy of  
Sciences, 25 Belsihuanxilu, Haidian, Beijing 100080, China

FEATURES  
source 1..1394  
/organism="Acanthopagrus schlegelii"  
/mol\_type="mRNA"  
/db\_xref="taxon:72011"  
1..>1394  
/gene="f1b"  
/note="affib"  
1..>1394  
/gene="f1b"  
/note="Orphan nuclear receptor"  
/product="FTZ-F1"  
/codon\_start=1  
/protein\_id="AA575792.1"  
/db\_xref="GI:45686255"  
/translation="MLGDAQGVTLKMEYTYDDLEELFCVCGDKVSGHYGLITCE  
SKGFFQRTVQNNKRYTCAENQCKIDTQKRCPCFRCFOKLSVGMLEAVRADRM  
GGRNFGFMYNDKQKALIRSNQFKLSVAPAPSLQDTYGFGLHSPLTI  
SKLLSPRSSITPTDYASLYGPPSLGMAQSHVPLTQYOTYFPGRAIKAECPDY  
TSSPSLTGYPPDMYSPASPOPSLPPIVLELRCDOSGARLDKPSFSLMCMAD  
QTLRSIVEMARSCIFFKELRVGDQKLIHNCWSEILVIDHI FROYOHQDESIILVTC  
QEVLSFTLSQSEANTLSLVGGGLAVRLALQVDRREVACLKFLLFNPVKLLED  
QAFEGVQEOVNGALLLEYLTSTYPOFORKFSQVLRVLRSLSTQARDYLCYMLSG  
EVPNNLLIEMLHA"

ORIGIN  
Alignment Scores:  
Pred. No.: 1.88 Length: 1394

Score: 10.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 3.64%  
DB: 5  
Matches: 10  
Conservative: 0  
Mismatch: 0  
Indels: 0  
Gaps: 0

US-10-069-353A-8 (1-275) x AY491379 (1-1394)

QY 244 ProAlaValAlaGlyTrpAlaAlaVal 253  
DB 742 CACGGCTGCGAGCTGCGCGCTGCGGCTG 713

## RESULT 13

LOCUS A60304 6085 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 4 from Patent WO9708323.  
ACCESSION A60304  
VERSION A60304.1 GI:3715279  
KEYWORDS  
SOURCE unidentified  
ORGANISM unidentified

REFERENCE 1 Schupp,T., Engel,N., Bietenhader,J., Toupet,C. and Pospiech,A.  
AUTHORS STAUFOSPORIN BIOSYNTHESIS GENE CLUSTERS  
TITLE Patent: WO 9708323-A 4 06-MAR-1997;  
JOURNAL CIBA GEIGY AG (CH)

## FEATURES

source Location/Qualifiers  
1..6085  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"  
misc\_RNA 378..1665  
/function="ORF"  
misc\_RNA 1747..2553  
/function="ORF"  
misc\_RNA 2593..4011  
/function="ORF"  
misc\_RNA 4013..4999  
/function="ORF"  
misc\_RNA 5071..6085  
/function="ORF"

## ORIGIN

## Alignment Scores:

Pred. No.: 7.12 Length: 6085  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.64% Indels: 0  
DB: 6 Gaps: 0

US-10-069-353A-8 (1-275) x A60304 (1-6085)

QY 51 AspArgLeuThrAspLeuValAlaGluArg 60  
DB 1894 GACCGCTCACCAATCTGTCGCGAGCGC 1923

## RESULT 14

LOCUS ARI44763 6085 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 4 from patent US 6210935.  
ACCESSION ARI44763  
VERSION ARI44763.1 GI:15106630  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unidentified.

REFERENCE 1 (bases 1 to 6085)  
AUTHORS Schupp,T., Engel,N., Bietenhader,J., Toupet,C. and Pospiech,A.  
TITLE Stauposporin biosynthesis gene clusters  
JOURNAL Patent: US 6210935-A 4 03-APR-2001;  
FEATURES Location/Qualifiers

source 1..6085  
/organism="unknown"  
/mol\_type="unassigned DNA"

## ORIGIN

Alignment Scores:  
Pred. No.: 7.12 Length: 6085  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.64% Indels: 0  
DB: 6 Gaps: 0

US-10-069-353A-8 (1-275) x ARI44763 (1-6085)

QY 51 AspArgLeuThrAspLeuValAlaGluArg 60  
DB 1894 GACCGCTCACCAATCTGTCGCGAGCGC 1923

## RESULT 15

LOCUS AE005764 10802 bp DNA linear BCT 12-JUN-2002  
DEFINITION Caulobacter crescentus CB15 section 90 of 359 of the complete genome.  
ACCESSION AE005764  
VERSION AE005764.1 GI:13422130  
KEYWORDS

SOURCE Caulobacter crescentus CB15  
ORGANISM Caulobacter crescentus CB15

REFERENCE 1 (bases 1 to 10802)  
AUTHORS Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E., Eisen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R., Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B., Debroy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L., Haft,D.H., Kolonay,J.F., Smit,J., Craven,M.B., Khouri,H., Shetty,J., Berry,K., Uterback,T., Tran,K., Wolf,A., Vamathevan,J., Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and Fraser,C.M.

Complete genome sequence of Caulobacter crescentus Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)  
JOURNAL MEDLINE 21173698  
PUBMED 11259647

REFERENCE 2 (bases 1 to 10802)  
AUTHORS Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R., Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B., Laub,M.T., Debroy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L., Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J., Berry,K., Uterback,T., Tran,K., Wolf,A., Vamathevan,J., Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and Fraser,C.M.

Direct Submission  
Submitted (31-JAN-2001) The Institute for Genomic Research, 9712  
JOURNAL Medical Center Dr. Rockville, MD 20850, USA  
TITLE Location/Qualifiers

FEATURES  
source 1..10802  
/organism="Caulobacter crescentus CB15"  
/mol\_type="genomic DNA"  
/strain="CB15"  
/db\_xref="taxon:190650"  
complement(86..262)  
/gene="CC0870"  
complement(86..262)  
/gene="CC0870"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK22855.1"  
/db\_xref="GI:13422131"  
/translation="MAFKRAALGLMLVAGAAQAAVTPAPVCGPLARFDKLPAS

```
gene PAPDGSVWVKTLIM"
CDS complement(341. .925)
/gene="CC0871"
complement(341. .925)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK22856.1"
/db_xref="GI:13422132"
/translation="MLIGRAASLPPPIPMNLPMPLILLYKDDIADIALVAGYRG
ETAAQQTSESELGGQRTDEASLRADLAQGSTLLTFRDAMKEPFCAMITPAD
DWYVMGITVETPTROGGMGRAVLETCAYARAGKVRMTVTIVSRDALIMYBERG
VYLGTGTOPFPYDDDRFGLPQRDDLAFFVMEKGL"
664. .1761
/gene="CC0872"
964. .1761
/gene="CC0872"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK22857.1"
/db_xref="GI:13422133"
/translation="MPDLRSRPMVRAMQARIVSLYRHPYKGFPTPERLERADLAAGCF
PCDRLAYAVEDSGSPDPAAPAHISKMKFTVLAKIPVARARTAYDEKSGVLTARAGR
PPAGDLROASGRAGREALLDLIDGAGCPKVIIEGPAHFMDSKSGVLTAVLAS
VADLSAKVRDIDLPLRFKRNLYYEGPANAENDLTGQTLISGAKAAVAKPIVRCAAT
HVDPTSERDIDLVKALFDHGHMFPGIYLVNTEGAVREGDVAAAG"
complement(1762. .2052)
/gene="CC0873"
complement(1762. .2052)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK22858.1"
/db_xref="GI:13422134"
/translation="MKPYRLSRRAKADLDITYSHORWGVQADYARELOATTEMI
AHPGWDGQDENLRAGYRCASGHVVFVGVREIILRVLHQSNNARHLG"
complement(2049. .2315)
/gene="CC0874"
complement(2049. .2315)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK22859.1"
/db_xref="GI:13422135"
/translation="MAKSKTSVVLGDHQAFTDSQVADGRYSASAEVIRAGLRLLIEN
EKLAALRALALEGESGFIEDFDFAITEERSASAPQGFHEE"
2305. .3258
/gene="CC0875"
2305. .3258
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="hemk family protein"
/protein_id="AAK22860.1"
/db_xref="GI:13422136"
/translation="MLAMADLYQNLVPHSISPPGRRRAKAPMTLTLVKAWTAADRL
KAGIDQPSIDARMLLEVAGVTRTEIVDPYRELASQIATLNDYLERARREPVSH
IIKRGKFWKILQVNNKNTVTPRPETEVIVDEVLKAPPEBMATSMGLGVSGSTILLAV
IAERPAKGLIGIDASEALAVARENANLINTNRALLHGDWTGTGSGSPLVVSNP
PVTPEVDTLEBEVRIHEPRILADGGPGLAAVRIAPETIIRVILKPGSLFAVELGID
OSQAVBALPRAAGATEVRIVKDLSTHDRVVLGVKAPLESPPA"
3517. .4629
/gene="CC0876"
3517. .4629
```

```
gene /gene="CC0876"
CDS /note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK22861.1"
/db_xref="GI:13422137"
/translation="WRDFKMGKROGRUNRGNGGKPOOHANRAFDNSNGPQVYRG
AAOSYKEYOOLARPASSSGDRLVAENYLOHAHYERFLRALIOPRPVADIVGKVYA
AYEIDFEAPPEPEVEVQPAQGAQEGEGENGERPRDRPENRRNRDNRDRNRENRD
KDRNDDNRNDSRDRWRDDDRPRDRPRDRPRDRPRDRPRDRPRDRPRDRPRDR
DRFRDRDRPRDRPRDRPAABEQPQEGEFRRSRRRRRNRDRDRDWOFGRRDP
MAVIEPEASPLTAEBAPAGEGSPMLGQGAVSHAPAFYGRARAPAEASAPAEAA
PASDGECEVRRRRARPSFEAAGABEGSGDEG"
4797. .5552
/gene="CC0877"
4797. .5552
/gene="CC0877"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK22862.1"
/db_xref="GI:13422138"
/translation="WTLPRLLIPLILAAACAGGCAYNPDIQRDQLIIVNDNLARDG
EKAMVDALRTSHVSKKEPKNARVOAGRVVAAAGLAGRPWEAVFLEAPAFVLPQ
GVYGVATVLLSLVENDDLAVIGHGAGVVVARHARAEVSOETTSKVLLGAKAAG
TEFGULADHDEBDARVYGLLIPFSKQLEMDKLGVDVMQAGIRPREAVVLMRMQA
LDGKSGGDLDSIHPSDAVRIQELERYIASKGM"
5651. .8230
/gene="CC0878"
5651. .8230
/gene="CC0878"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="ATP-dependent Clp protease, ATP-binding subunit
ClpB"
/protein_id="AAK22863.1"
/db_xref="GI:13422139"
/translation="WNIDLVSDRAKQAVOSALALAHQQAPAEHILKYLEBKDG
LSRLAIQAGSRPDQLDGVEVTLAKTRPVDAAGQALYMKDPTARVFEAKSAAAG
DAFVTERLLIAlAKEGGEAAKLFKEAGVSAQSLETANANRKGRTADSAABEYEA
LKRYARLDLTAARDDKLDPIVGRDEIRRTIOVLSRTKNNPVLIGEGVGKTAIVEG
LAIRIVNGDVPESIKDKKLLSLDMSGLIAGAKYGEPEERLKAVIGEVTAEGSTIIP
IDENATVYAGKSGDSAMSLKRALRAGELHCYGAATTDEYRKHYKDAALARRQ
PVFVEPTVEDIVSTLIRGLKEKYEYHGVRLSDSAIVAATLSNRYIADRFLPDKAID
LVDEASRVRMQIDSKEPELDEIDRLVQLTIERALSKEVDASKQRLNILEVIDD
LOFSDEMTARMAKEKEKVGGAQAREALDRADLANAQAQGFARAQIOYGEIIPA
LEBRILAEAGDTQALTPREVDAEOLIAVVRNMGVPVEKMLEGRELKSLAMEDELAR
RVGQDEALAEVAVRVARAGIOPDSKPSGSPFLIGPTGVGKTELKSLAEPLPDE
AAITMDSEYMEKHSVSRLLGAPPGYVYDEGALTAIRRPYQVILPBEIEKAP
DVFNVLQVLDLDRITDQGRVDPERNLTIIMTSLVGEVIALSQDSQDAAVALAPMVM
NTVRGHFPERFLNRIDELILFRKLSRNHMGIOVILQGRVKLLADRRMALALAEAL
NMLADKGYDPPYAGAPLKRVIQKEIVPIAKKLAGEIEDGVIAGVTQQLSPGKA
VLQ"
8231. .8782
/gene="CC0879"
8231. .8782
/gene="CC0879"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK22864.1"
/db_xref="GI:13422140"
/translation="WTSLSISPCGGMPPRRKSGEGLKTISAATGLSAGRGFRSRDP
SSVAFSDFDSKGRGAFRPPPALSPRMATKTKLLITPEPIVDGVDYVIRHLPLPO
WSYDSEBHGGVIAARHKGOREALIEPPPKRKFSAPLAKPSQIPRIYAKALASA
IEAGWEPUSRGKAVIIVDEGA"
complement(8792. .9070)
/gene="CC0880"
complement(8792. .9070)
```



```
/gene="CC0880"  
/note="identified by Glimmer2; putative"
```

/note="identified by Glimmer2; putative"

Alignment Scores:

Pred. No.:	12	length:	10802
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.64	Indels:	0
DB:	1	Gaps:	0

Score:	10.00	Matches:	10
--------	-------	----------	----

Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

```
Best local similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
```

DB:	1	Gaps:	0
-----	---	-------	---

TUC 10 000 3532 8 (1 27E) ~ 2E00E724 (1 10003)

US-10-069-353A-8 (1-275) X AE005764 (1-10802)

Db 3076 GAGATTCTGCGGCTCTCAGCCCGGCGT 3105

Search completed: August 30, 2005, 00:04:50  
Job time : 4806 secs

Job time : 4806 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 29, 2005, 20:03:44 ; Search time 611 Seconds  
(without alignments)  
2664.368 Million cell updates/sec

Title: US-10-069-353a-8

Perfect score: 275  
Sequence: 1 VLPGAPISQVGVQMYDLVT.....YEKIAHMGVAILTARKPVG 275

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 segs, 2959870667 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8776198

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh  
-O=/cgm2\_1/uspto.spool/US10069353.rnaut 26082005 172151 15042/app.query.fasta\_1.455  
-DB=N\_Geneseq 16Dec04 -OPMT=fastap -SUFFIX=ring -MIMATCH=0.1 -LOOFCI=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USR=US10069353 @CGN 1.1 470 @rnaut 26082005 172151 15042 -NCPU=6 -ICPU=3  
-NO MAMP -IARGEBLERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEXTIMEOUT=120 -MAMP\_TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6  
-Fgapext=7 -Ygapop=60 -Ygapext=60 -DELOP=6 -DELEXT=7

Database: N\_Geneseq 16Dec04:\*

1: geneseqn1980s:.\*  
2: geneseqn1980s:.\*  
3: geneseqn2000s:.\*  
4: geneseqn2001as:.\*  
5: geneseqn2001bs:.\*  
6: geneseqn2002as:.\*  
7: geneseqn2002bs:.\*  
8: geneseqn2003as:.\*  
9: geneseqn2003bs:.\*  
10: geneseqn2003cs:.\*  
11: geneseqn2003ds:.\*  
12: geneseqn2004as:.\*  
13: geneseqn2004bs:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	275	100.0	828	4	AAf88318
2	275	100.0	45624	4	AAf88315
3	275	100.0	50000	4	AAf88312
4	275	100.0	80161	2	AA221501
5	37	13.5	36538	10	ABV75558

Result ID	Score	Query Match	Length	DB ID	Description
1	351	9	AC122474	9	AC122474 DNA clone
2	397	9	AC122489	9	AC122489 DNA clone
3	406	9	AC122477	9	AC122477 DNA clone
4	439	9	AC122483	9	AC122483 DNA clone
5	552	9	AC122491	9	AC122491 DNA clone
6	572	9	AC122487	9	AC122487 DNA clone
7	607	9	AC122468	9	AC122468 DNA clone
8	617	9	AC122454	9	AC122454 DNA clone
9	632	9	AC122486	9	AC122486 DNA clone
10	633	9	AC122484	9	AC122484 DNA clone
11	642	9	AC122476	9	AC122476 DNA clone
12	651	9	AC122473	9	AC122473 DNA clone
13	654	9	AC122488	9	AC122488 DNA clone
14	675	9	AC122475	9	AC122475 DNA clone
15	685	9	AC122482	9	AC122482 DNA clone
16	703	9	AC122478	9	AC122478 DNA clone
17	703	9	AC122458	9	AC122458 DNA clone
18	705	9	AC122490	9	AC122490 DNA clone
19	708	9	AC122485	9	AC122485 DNA clone
20	719	9	AC122472	9	AC122472 DNA clone
21	721	9	AC122479	9	AC122479 DNA clone
22	726	9	AC122480	9	AC122480 DNA clone
23	738	8	ABX56032	8	ABX56032 M. echino
24	6085	2	AAf70153	2	AAf70153 S. longisp
25	759	8	ACF73542	8	ACF73542 Staphyloc
26	11466	2	AAf74755	2	AAf74755 Staphyloc
27	47745	11	ACN44160	11	ACN44160 Mouse gen
28	64492	12	ADK16023	12	ADK16023 Streptomy
29	87	8	AB209681	8	AB209681 Human o1
30	87	10	ABN71244	10	ABN71244 Streptoco
31	213	6	ABN71244	6	ABN71244 Streptoco
32	270	6	ABE65692	6	ABE65692 Lung canc
33	270	6	ABK64644	6	ABK64644 Human ben
34	270	6	ABV77957	6	ABV77957 Hypoxia-r
35	349	5	ABV07585	5	ABV07585 Human pro
36	382	10	ADG38142	10	ADG38142 Aspergill
37	405	5	ABV37517	5	ABV37517 Human pro
38	414	11	ABD01730	11	ABD01730 Pseudomon
39	441	10	ABX07019	10	ABX07019 S. pneumo

#### ALIGNMENTS

RESULT 1  
ID AAF88318 standard; DNA; 828 BP.  
AC AAF88318;  
XX  
DT 28-AUG-2001 (first entry)  
XX  
DE S. spinoza DNA fragment encoding ORF1, SEQ ID 7.  
XX  
KW Forosamine; trimethylthiamine; polyketide synthase; biosynthesis;  
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;  
KW microtubule; insecticide; O-methyltransferase; ds.  
XX  
OS Saccharopolyspora spinoza.  
XX  
PN DE19957268-A1.  
XX  
PD 08-MAR-2001.  
XX  
PF 29-NOV-1999; 99DE-01057268.  
XX  
PR 27-AUG-1999; 99DE-01040596.  
XX  
PA (FARB ) BAYER AG.  
XX  
PI Ebertz G, Moehrle V, Froede R, Velten R, Salas JA;  
XX  
DR WPI; 2001-267102/28.

DR P-PSDB; AAB70948.  
 XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for  
 PT recombinant production of insecticidal spinosyns and their derivatives.  
 XX  
 PS Claim 7a; Page 102-104; 354pp; German.  
 XX  
 CC This invention describes a novel method nucleic acid (I) and its encoded  
 CC polypeptide (II) containing at least one region that encodes an enzymatic  
 CC activity involved in biosynthesis of spinosyns. (I) are used (i) to  
 CC identify, inactivate or modulate genes involved in the biosynthesis of  
 CC (ii); (iii) to generate a library of polypeptide synthetases; (iii) for  
 CC adding forosamine or trimethylrhamsome to a spinosyn or polypeptide  
 CC aglycone; and (iv) for recombinant production of the corresponding  
 CC enzymes, which are used for production of (II), their precursors or  
 CC derivatives, including production of transgenic plants that express (II)  
 CC and thus have increased resistance to insects. (i) are also useful as  
 CC markers for sequencing the Saccharopolyspora spinosa genome. (ii) are  
 CC macrolides with insecticidal, but not antibacterial, activity, and can  
 CC also be used to raise specific antibodies, useful for identifying  
 CC expression clones in a gene bank. Cells transformed with (I) may produce  
 CC (II) at significantly increased levels or produce new derivatives of  
 CC (II). This sequence, ORF 1, encodes an S. spinosa O-methyltransferase  
 CC protein  
 XX  
 SQ Sequence 828 BP; 140 A; 262 C; 285 G; 141 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 3,88e-265 Length: 828  
 Score: 275.00 Matches: 275  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 4 Gaps: 0  
 US-10-069-353a-8 (1-275) x AAF88318 (1-828)  
 QY 1 ValLeuProGlyValAlaProThrSerGlnGlnValAlaGlyGlnMetTyrAspLeuValThr 20  
 DB 1 GTGTTCGCAAGGTGGCGCAACCAATCGACAGGTGGAGTGGATGACCTGGTACG 60  
 QY 21 ProLeuLeuAnsSerValAlaGlyGlyProCysAlaIleHisIshGlyTyrTrpGluAsn 40  
 DB 61 CCGTTCGTAACCTGGTGGGGGGGGCCCTGGCCCATCCACCGGCTACTGGGAAAC 120  
 QY 41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60  
 DB 121 GACGGGGGGGGCTTCCTGGCAGCAGCGCCGCGACCGGCTCAACGCTTGGCCGAAACGG 180  
 QY 61 ThrValLeuAspGlyValAlaArgLeuLeuAspValAlaGlyCysGlyThrGlyGlnProAla 80  
 DB 181 ACCGTGCTCGATGGCGGGGCTTCGACTGCTGATGTGGGGTGGGTACCGGAAACGACGGC 240  
 QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100  
 DB 241 CTGGGGCTGGGGCGGCGCAACACGGCATCCAGATCAACGGGCAATCCCTGACGCAAGTCAA 300  
 QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120  
 DB 301 GTGGCCCTCGCGCTGATTGGCGCAGCGGAACGCGACCTAAGCCACCGGGTGGACTTCTCG 360  
 QY 121 CysValAlaAspAlaMetSerLeuProTyrProAspAsnAlaIleAspAlaAlaTPAlaMet 140  
 DB 361 TGGCTCATGCAATGCCCTGCGCTACCGGCAATCTTTCGAGCGGCGCTGGGCGCATG 420  
 QY 141 GlnSerLeuLeuGluMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160  
 DB 421 CAGTCCGTGTTGGAGATGTCGGAACCGGACCGTCCATCCGGGAAATCTTCAGATCTC 480  
 QY 161 LysProGlyGlyIleLeuGlyValAlaThrGluValAlaLysArgGluAlaGlyGlyMet 180  
 DB 481 AAACCCCGTGGCATCTTCGCGGTCCACGAGAGTGTCTAAACGAGAACGGGGCGGGGATG 540

QY 181 ProValSerGlyAspArgTrpProThrGlyLeuArgIleCysLeuAlaGluGlnLeu 200  
 DB 541 CCGGTGTCGGGAGACAGGTGGCCGACCGGCTTCGATTCGCTGCTGAGCAATCTTG 600  
 QY 201 GluSerLeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220  
 DB 601 GAATCCGTGCTGCACACGGGGTTCGAGATCTCGATTGGAGAGACCTGTCTCGAGAGAC 660  
 QY 221 ArgTyrPheMetProGlnPheAlaGluGluLeuAlaAlaHisGlnHisGlyIleAlaAsp 240  
 DB 661 CGGTACTTCATGCGCAGATTCGCCGAAGAGCTGCGCCACACGACGAGATCGGGGAC 720  
 QY 241 ArgTyrGlyProAlaValAlaGlyTyrAlaAlaValCysAspTyrGlyLysTyrAla 260  
 DB 721 AGGTACGGGGCGGCTGTGCGGCTGGCCCGCGGCTGCGATTATGAAATATGCC 780  
 QY 261 HisAspMetGlyTyrAlaIleLeuThrAlaArgLysProValGly 275  
 DB 781 CACGACATGGGCTATGCGATTCTGACGGCGGGAACCGGTCCGC 825  
 RESULT 2  
 AAF88315/c  
 ID AAF88315 standard; DNA; 45624 BP.  
 XX  
 AC AAF88315;  
 XX  
 DT 28-AUG-2001 (first entry)  
 XX  
 DE S. spinosa DNA fragment SEQ ID 4.  
 XX  
 KW Forosamine; trimethylrhamsome; polyketide synthase; biosynthesis;  
 KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;  
 KW macrolide; insecticidal; ds.  
 XX  
 OS Saccharopolyspora spinosa.  
 XX  
 PN DE19957268-A1.  
 XX  
 PD 08-MAR-2001.  
 XX  
 PF 29-NOV-1999; 99DE-01057268.  
 XX  
 PR 27-AUG-1999; 99DE-01040596.  
 XX  
 PA (FARB ) BAYER AG.  
 XX  
 PI Eberz G, Moehle V, Froede R, Valten R, Salas JA;  
 XX  
 DR WPI; 2001-267102/28.  
 XX  
 PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for  
 PT recombinant production of insecticidal spinosyns and their derivatives.  
 XX  
 PS Claim 7; Page 58-74; 354pp; German.  
 XX  
 CC This invention describes a novel method nucleic acid (I) and its encoded  
 CC polypeptide (II) containing at least one region that encodes an enzymatic  
 CC activity involved in biosynthesis of spinosyns. (I) are used (i) to  
 CC identify, inactivate or modulate genes involved in the biosynthesis of  
 CC (ii); (iii) to generate a library of polypeptide synthetases; (iii) for  
 CC adding forosamine or trimethylrhamsome to a spinosyn or polypeptide  
 CC aglycone; and (iv) for recombinant production of the corresponding  
 CC enzymes, which are used for production of (II), their precursors or  
 CC derivatives, including production of transgenic plants that express (II)  
 CC and thus have increased resistance to insects. (i) are also useful as  
 CC markers for sequencing the Saccharopolyspora spinosa genome. (ii) are  
 CC macrolides with insecticidal, but not antibacterial, activity, and can  
 CC also be used to raise specific antibodies, useful for identifying  
 CC expression clones in a gene bank. Cells transformed with (I) may produce  
 CC (II) at significantly increased levels or produce new derivatives of  
 CC (II). This sequence represents a genomic DNA fragment of the S. spinosa  
 CC genome which contains the coding regions for proteins involved in  
 CC forosamine and trimethylrhamsome biosynthesis

XX Sequence 45624 BP; 7933 A; 14913 C; 14940 G; 7838 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	1,74e-263	Length:	45624
Score:	275.00	Matches:	275
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	4	Gaps:	0

US-10-069-353A-8 (1-275) x AAF88315 (1-45624)

```

QY 1 valleuproglygylalaprothserglnlnvalgylgmetyrarspleuvalthr 20
DB 828 GTGTTGCCAGTGGCGCACCAATCGCAGAGTGGGAGATGATATACCTGGTCAAG 769
QY 21 Prolenleuanservai1aagiylgyprocysalalehihiagiyltyrtrpjuu 40
DB 768 CCGTTCCTGAACCTGGTCCGGGGGGCCCTCGCATCCACACCGGCTACTGGAGAAC 709
QY 41 AapglYArgAlaserTrpIngnlnalaaaspargleuthraspleuvalaagiua 60
DB 708 GACGGGCGGGCTTCCTCGCAGCAGCGCCGCGACCGGCTCAGCAGCTTGTCCGCAAG 649
QY 61 Thrvalleuaspolygylvalargleuaspvaliglycyaglythrlyglnproa 80
DB 648 ACCGTCTCGATGGCGCGCTTCGACTGCTGATGGGGGTGGCGGTACCGGACCAACGCG 589
QY 81 leuargvalalargaphenmaliegnlnlethgyltlethralserglnvalgin 100
DB 588 CTGGCGGTGGCGGCGCACACCGGATCCAGATCCAGGATCCCTCAGCCAGGTCGAA 529
QY 101 valalalalalalalaspysalaaargluarglyleuserhisargvalaspheser 120
DB 528 GTGGCCTGGCGGCTTCCTCGCAGCAGCGGACCGGAGCTAAGCAGCGGGGTGACTTCTG 469
QY 121 Cysvalaspalawetserleuprotyrproaspasnalaphaspalalatrpalamet 140
DB 468 TGGGTGATGCGATGTCCTCGCGGTACCCGCGACATGCTTTCACCCCGCTGGGCGCATG 409
QY 141 Glnserleuengluwetsergyluproaspargalalalargluilleuargvalleu 160
DB 408 CAGTCTCTGTGAGATGTCGAGACCGGACCGTGCATCCGGAHAATCTTCAGATACCTC 349
QY 161 lvsprgylgylleuenglyvalthrlnuvalvallysaargyluaagiylgylmet 180
DB 348 AAACCGGCGGCTCTCTCGGCGTCACCGAGTCTCTCAACGAGAACCGGGCGGGGATG 289
QY 181 Provalsergylaspargyltrporthgylleuargylecysleuvalagluinleu 200
DB 288 CCGGTGTCGGGAGAGGTGGCGGACCGGCGCTTCGATCGCTGCTGGCTGCAACTCTCTG 229
QY 201 Gluserleuarglalaagiylphegylulleuaspfrglunaspvalaserargthr 220
DB 228 GAATCGTGGCGAGCGGGGTCGAGATCTCGATTGGGAGAGCGTGTGAGAGACC 169
QY 221 Argtyrphemetproginphealagluinleualalalhisglnhisgyllealasp 240
DB 168 CCGTACTTCATGCGCGAGTTCGCGGAAGACTCGCTCGCAGCAGCGGAGTCGGGAGC 109
QY 241 Argtyrlyproalavalalaglytrpalaalalavalalalaspaspyrtyrlyu 260
DB 108 AGGTACGGGCGCGGTGTCGCGGGTGGGCGCGGCTGTGCAATATGAAATATGCC 49
QY 261 Hisaspmetgyltyrallaleuthrlyalarglyserprovalgyl 275
DB 48 CACGACATGAGGCTATGCAATTCGACGCGCGGAGACCGGATCGGC 4

```

RESULT 3  
AAF88312  
ID AAF88312 standard; DNA; 50000 BP.  
XX

```

AC AAF88312;
XX 28-AUG-2001 (first entry)
XX S. spinosa DNA fragment SEQ ID 1.
XX
XX Forosamine; trimethylrhamsone; polyketide synthase; biosynthesis;
XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;
XX macrolide; insecticidal; de.
XX Saccharopolyspora spinosa.
XX DE19957268-A1.
XX 08-MAR-2001.
XX 29-NOV-1999; 99DE-01057268.
XX 27-AUG-1999; 99DE-01040596.
XX (FARB ) BAYER AG.
XX Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX WPI; 2001-267102/28.
XX
XX Claim 7, Page 14-31; 354p; German.
XX
XX This invention describes a novel method nucleic acid (I) and its encoded
XX polypeptide (II) containing at least one region that encodes an enzymatic
XX activity involved in biosynthesis of spinosyns. (I) are used (i) to
XX identify, inactivate or modulate genes involved in the biosynthesis of
XX (II); (ii) to generate a library of polyketide synthases; (iii) for
XX adding forosamine or trimethylrhamsone to a spinosyn or polyketide
XX aglycone; and (iv) for recombinant production of the corresponding
XX enzymes, which are used for production of (II), their precursors or
XX derivatives, including production of transgenic plants that express (II)
XX and thus have increased resistance to insects. (I) are also useful as
XX markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
XX macrolides with insecticidal, but not antibacterial, activity, and can
XX also be used to raise specific antibodies, useful for identifying
XX expression clones in a gene bank. Cells transformed with (II) may produce
XX (II) at significantly increased levels or produce new derivatives of
XX (II). This sequence represents a genomic DNA fragment of the S. spinosa
XX genome which contains the coding regions for proteins involved in
XX forosamine, trimethylrhamsone and polyketide synthase biosynthesis
XX
XX SQ Sequence 50000 BP; 8484 A; 16290 C; 16535 G; 8691 T; 0 U; 0 Other;

```

## Alignment Scores:

Pred. No.:	1,89e-263	Length:	50000
Score:	275.00	Matches:	275
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	4	Gaps:	0

US-10-069-353A-8 (1-275) x AAF88312 (1-50000)

```

QY 1 valleuproglygylalaprothserglnlnvalgylgmetyrarspleuvalthr 20
DB 44797 GTGTTGCCAGTGGCGCACCAATCGCAGAGTGGGAGATGATATACCTGGTCAAG 44856
QY 21 Prolenleuanservai1aagiylgyprocysalalehihiagiyltyrtrpjuu 40
DB 44857 CCGTTCCTGAACCTGGTCCGGGGGGCCCTCGCATCCACACCGGCTACTGGAGAAC 44916
QY 41 AapglYArgAlaserTrpIngnlnalaaaspargleuthraspleuvalaagiua 60
DB 44917 GACGGGCGGGCTTCCTCGCAGCAGCGCCGCGACCGGCTCAGCAGCTTGTCCGCAAGC 44976

```

QY		61	ThrValIleuAspGlyValAlaArgLeuIleuAspValGlyCysGlyTyrThrGlyInProAla	80		FT	CDS		complement(3416..4165)
FT						FT		/tag= c	
Db		44977	ACCGTCGTCAAGGGGGGGCTTGACTGCTCGATGGGGGTGGTAACGGAACACCACGC	45036		FT		/product= "spnS"	
FT						FT		/note= "Spinosyn biosynthesis protein S"	
QY		81	IeuAArgValAlaAraAspAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln	100		FT	CDS	complement(4168..5325)	
FT						FT		/tag= d	
Db		45097	CTGCCGCGTCGGCGGACAACCCGATCCAGATCACCGCATCACGTCAGCCAGGTGCAC	45096		FT		/product= "spnR"	
FT						FT	CDS	/note= "Spinosyn biosynthesis protein R"	
QY		101	ValAlaIleAlaAlaAspCysAlaAargLuiAargLyleuSerHisAargValAspPheSer	120		FT		/product= "spnO"	
FT						FT		/note= "Spinosyn biosynthesis protein Q"	
Db		45097	GTTGGCCATCGCCGCTGATTGCGCACCGCAACCGCATTAAGCCACCGGTTGACTTCTCG	45156		FT	CDS	/tag= e	
FT						FT		/note= "Spinosyn biosynthesis protein Q"	
QY		121	CysValAspAlaMetSerIleuProTyrProAspAsnAlaPheAspAlaAlaITrpaIleMet	140		FT		/product= "spnP"	
FT						FT		/note= "Spinosyn biosynthesis protein P"	
Db		45157	TTCGTGATGCCATGTCTCTGCGTAACCGGAACAAATGCTTTGACGCCGCTGGGCCATG	45216		FT	CDS	/tag= f	
FT						FT		/product= "spnQ"	
QY		141	GlnSerIleuLeuGluMetSerGluProAspAralaIeAargLuiIleuAryValIleu	160		FT		/tag= g	
FT						FT		/product= "spnO"	
Db		45217	CAGTCGCTGTGGAGATGTCGCAACCGGACCTGCATCCGGAGAATCCTTCGACTACTC	45276		FT		/note= "Spinosyn biosynthesis protein O"	
FT						FT	CDS	/tag= h	
QY		161	LysProGlyValIleLeuGlyValThrGluValValLysAargLuiAlaGlyGlyMet	180		FT		/product= "spnN"	
FT						FT		/note= "Spinosyn biosynthesis protein N"	
Db		45277	AAACCCGGGGGATCTCGGCGTCAACGAGTGTCAAACGAACGAGCGCGCGGGATG	45336		FT	CDS	/tag= i	
FT						FT		/product= "spnM"	
QY		181	ProValSerGlyAspAargTrpProThrGlyLeuAglileCysLeuAlaGluGlnIleu	200		FT		/note= "Spinosyn biosynthesis protein M"	
FT						FT	CDS	/tag= j	
Db		45337	CCGGTGTCCGGGGAACAGGTGGCGAACCGGCTTCGGATCTGCTGGAGCAATTCTTG	45396		FT		/product= "spnL"	
FT						FT		/note= "Spinosyn biosynthesis protein L"	
QY		201	GluSerIleuAargAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerAargThr	220		FT		/tag= k	
FT						FT		/product= "spnK"	
Db		45397	GAATGCGCGCGGACGCGGGTTCAGATCTTCGATTGGAGAGACGTGTCTGCGAGAAC	45456		FT	CDS	/tag= l	
FT						FT		/note= "Spinosyn biosynthesis protein K"	
QY		221	AArgTyrPheMetProGluPheAlaGluGluLeuAlaAlaHisGlnHisGlyYlleAlaAsp	240		FT		/product= "spnJ"	
FT						FT		/note= "Spinosyn biosynthesis protein J"	
Db		45457	CGGTACTTCATGCCCAAGTCTGCCCAAGAGCTCGCTGGCACACGACGCGATCCGGAC	45516		FT	CDS	/tag= m	
FT						FT		/product= "spnI"	
QY		241	AArgTyrGlyProAlaValAlaGlyTyrTrpaAlaAlaValCysAspTyrGluLysTyrTrla	260		FT		/note= "Spinosyn biosynthesis protein I"	
FT						FT		/tag= n	
Db		45517	AGGTACGGGCGCGGTGCGCCGCGTGGCGCGCGGTCTGGATTATGAAATATATCCC	45576		FT	CDS	/tag= o	
FT						FT		/product= "spnH"	
QY		261	HIsAspMetGlyTyrAlaIleIleuThrAlaAargLysProValGly	275		FT		/note= "Spinosyn biosynthesis protein H"	
FT						FT	CDS	/tag= p	
Db		45577	CACGACATGGGCTATGCAATTCGACGGGCGGAACCGGTCGGC	45621		FT		/product= "spnF"	
FT						FT		/note= "Spinosyn biosynthesis protein F"	
RESULT 4									
XX		AAZ21501				FT		/tag= q	
AC		AAZ21501	standard; DNA; 80161 BP.			FT	CDS	/note= "Beta-ketosynthase domain (KS1) : part of the initiator module"	
XX		AAZ21501;				FT		/tag= s	
DT		01-DEC-1999	(first entry)			FT		/note= "Acyl transferase domain (AT1) : part of the initiator module"	
XX		DE	DNA fragment of Saccharopolyspora spinosa containing biosynthetic genes.			FT		/tag= t	
XX		XX	Spinosyn biosynthetic enzyme; open reading frame; ORF:			FT		/note= "Acyl carrier protein domain (ACP1) : part of the initiator module"	
KW			insecticidal microlides; arachnid; nematode; insect; polyketide;			FT	CDS	/tag= u	
KW			polyketide synthase; PKS; extender module; initiator module;			FT		/product= "spnA"	
KW			acyl transferase domain; AT; acyl carrier protein; ACP;			FT		/note= "Spinosyn biosynthesis protein A"	
KW			beta-ketosynthase domain; KS; KR; dehydrodase domain; DH;			FT		/transl_except= (Pos:26940..26942, aa:Pro)	
KW			enoyl reductase domain; ER; beta-ketoreductase; insecticide; ss.			FT		21126..22379	
XX		OS	Saccharopolyspora spinosa.			FT		/tag= x	
XX		Key	Location/Qualifiers			FT		/note= "Beta-ketosynthase domain (KS1) : part of the initiator module"	
FT		CDS	complement(1135..1971)			FT		/tag= y	
FT			/tag= a			FT		/note= "Acyl transferase domain (AT1) : part of the initiator module"	
FT			/product= "ORFL16"			FT		/note= "Acyl transferase domain (AT1) : part of the initiator module"	
FT			/note= "Protein involved in transcription control"			FT		2024..2791	
FT		CDS	/tag= b			FT		/tag= z	
FT			/product= "ORFL15"			FT		/note= "Keto acyl reductase"	
FT			/note= "Keto acyl reductase"			FT			

```

FT misc_feature 24102..25349
FT /tag= u
FT /note= "Beta-ketosynthase domain (KS1) : part of extender
FT module 1"
FT misc_feature 25683..26684
FT /tag= v
FT /note= "Acyl transferase domain (AT1) : part of extender
FT module 1"
FT misc_feature 27582..28121
FT /tag= w
FT /note= "Beta-ketoreductase domain (KR1) : part of extender
FT module 1"
FT misc_feature 28404..28649
FT /tag= x
FT /note= "Acyl carrier protein domain (ACP1) : part of
FT extender module 1"
FT misc_feature 29024..30295
FT /tag= z
FT /note= "Beta-ketosynthase domain (KS2) : part of extender
FT module 2"
FT misc_feature 30629..31621
FT /tag= aa
FT /note= "Acyl transferase domain (AT2) : part of extender
FT module 2"
FT misc_feature 31697..32254
FT /tag= ab
FT /note= "Dehydratase domain (DH2) : part of extender module
FT 2"
FT misc_feature 33035..34072
FT /tag= ac
FT /note= "Enoyl reductase domain (ER2) : part of extender
FT module 2"
FT misc_feature 34082..34621
FT /tag= ad
FT /note= "Beta-ketoreductase domain (KR2) : part of extender
FT module 2"
FT misc_feature 34886..30295
FT /tag= ae
FT /note= "Acyl carrier protein domain (ACP2) : part of
FT extender module 2"
FT CDS 35419..44931
FT /tag= af
FT /product= "spnc"
FT /note= "Spinosyn biosynthesis protein C"
FT 35518..36786
FT /tag= ag
FT /note= "Beta-ketosynthase domain (KS3) : part of extender
FT module 3"
FT misc_feature 37108..38097
FT /tag= ah
FT /note= "Acyl transferase domain (AT3) : part of extender
FT module 3"
FT CDS 38916..35374
FT /tag= y
FT /product= "spnc"
FT /note= "Spinosyn biosynthesis protein C"
FT 38992..39528
FT /tag= ai
FT /note= "Beta-ketoreductase domain (KR3) : part of extender
FT module 3"
FT misc_feature 39790..40035
FT /tag= aj
FT /note= "Acyl carrier protein domain (ACP3) : part of
FT extender module 3"
FT misc_feature 40102..41373
FT /tag= ak
FT /note= "Beta-ketosynthase domain (KS4) : part of extender
FT module 4"
FT misc_feature 41713..42705
FT /tag= al
FT /note= "Acyl transferase domain (AT4) : part of extender
FT module 4"
FT misc_feature 43615..44157

```

```

FT /tag= am
FT /note= "Beta-ketoreductase domain (KR4) : part of extender
FT module 4"
FT misc_feature 44431..44676
FT /tag= an
FT /note= "Acyl carrier protein domain (ACP4) : part of
FT extender module 4"
FT CDS 44966..59752
FT /tag= ao
FT /product= "spnd"
FT /note= "Spinosyn biosynthesis protein D"
FT 45077..46348
FT misc_feature 45077..46348
FT /tag= ap
FT /note= "Beta-ketosynthase domain (KS5) : part of extender
FT module 5"
FT misc_feature 46691..47674
FT /tag= aq
FT /note= "Acyl transferase domain (AT5) : part of extender
FT module 5"
FT misc_feature 47753..48310
FT /tag= ar
FT /note= "Dehydratase domain (DH5) : part of extender module
FT 5"
FT misc_feature 49226..49771
FT /tag= as
FT /note= "Beta-ketoreductase domain (KR5) : part of extender
FT module 5"
FT misc_feature 50009..50254

```

## Alignment Scores:

```

Pred. No.: 2,966-263 Length: 80161
Score: 275.00 Matches: 275
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

```

US-10-069-353a-8 (1-275) x AAZ21501 (1-80161)

```

QY 1 ValLeuProGlyValAlaProThSerGlnGlnValGlyMetTyrAspLeuValThr 20
DB 20168 GTGTTGCCAGTGGCGCACCAACATCGCAGCGAGTTGGGCGAGATGATGATCGTGCACG 20227
QY 21 ProLeuLeuAsnSerValAlaGlyIleProCysAlaIleHisGlyTyrTrpGluAsn 40
DB 20228 CCGTTCGTAACCTCGATCGCGGCGCGCCCTCGCATCCACCGAGCTACTGGAGAAC 20287
QY 41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
DB 20288 GACGGCGGAGCTTCTCTGCGACAGCGCCCGCACCGGCTCACCGACTTGTCCGACACGG 20347
QY 61 ThrValLeuAspGlyValValArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80
DB 20348 ACCGTCCTCGATGGCGCGCTTCGACTGCTCGATGCGGCGTGCAGTACCGAACACACGG 20407
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
DB 20408 CTGCGCGTCCGCGGACCAACCGATCCAGATCACCGGATCACCGTCCGACGAGTGCA 20467
QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
DB 20468 GTGGCCATCGCGCTGATGGCCGACGCGACCGAGCTPAAGCACCGGTGAGCTTCTCG 20527
QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
DB 20528 TCGCTGATGATGATGCTCTGCGTACCGGACCAATGCTTTGAGCGCCGCTCGGCGCAG 20587
QY 141 GlnSerLeuLeuGluMetSerGluProAspArgAlaIleArgGlnIleLeuArgValLeu 160
DB 20588 CAGTCGCTGTGAGATGTCACACCGGACCGTGCATCCGCGGAATCTTGAGTACTC 20647
QY 161 LysProGlyGlyIleLeuGlyValThrGluValValLysArgGluAlaGlyValGlyMet 180

```







```
XX
SQ Sequence 397 BP; 105 A; 92 C; 70 G; 130 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.1 Length: 397
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 9 Gaps: 0

US-10-069-353a-8 (1-275) x ACL22477 (1-397)

Qy 148 GIUProASPARGAlaIleArgGluIleLeu 157
Db 238 GAACCGATCGTGGATTGGAGATTTTA 209

RESULT 8
ACL22477/c
ID ACL22477 standard; DNA; 406 BP.
XX
AC ACL22477;
XX
DT 27-OCT-2003 (revised)
DT 17-OCT-2003 (first entry)
XX
DE DNA clone originating in barley containing SNP encoding sequence #12468.
XX
KM Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KW gene; ss.
XX
OS Hordeum vulgare; var. (cul.Haruna Nijo).
XX
PN WO2003057877-A1.
XX
PD 17-JUL-2003.
XX
PF 16-DEC-2002; 2002WO-IB005403.
XX
PR 20-DEC-2001; 2001JP-00387059.
XX
PR 20-DEC-2001; 2001JP-00387131.
XX
PR 20-DEC-2001; 2001JP-00403299.
XX
PR 20-DEC-2001; 2001JP-00403300.
XX
PR 27-SEP-2002; 2002JP-00327515.
XX
PA (UYN1-) UNIV JAPAN OKAYAMA.
XX
PI Sato K, Takeda K, Kohara Y;
XX
DR WPI; 2003-587127/55.
XX
PT Single nucleotide polymorphism sites in barley varieties and DNA
PT sequences containing them for analysis and identification of barley
PT varieties and production of barley transformants with desired
PT characteristics.
XX
PS Disclosure; SEQ ID XX; 284bp; Japanese.
XX
CC The present invention relates to oligonucleotide clones originating in
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
CC varieties, identification of particular varieties and genotype-phenotype
CC analysis, isolation of specific genes and creation of new varieties by
CC transformation of barley varieties with them and production of new barley
CC oligonucleotide clone DNA sequence featured in the specification. The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 406 BP; 108 A; 96 C; 74 G; 127 T; 0 U; 1 Other;
```

```
Alignment Scores:
Pred. No.: 2.15 Length: 406
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 9 Gaps: 0

US-10-069-353a-8 (1-275) x ACL22477 (1-406)

Qy 148 GIUProASPARGAlaIleArgGluIleLeu 157
Db 168 GAACCGATCGTGGATTGGAGATTTTA 139

RESULT 9
ACL22483/c
ID ACL22483 standard; DNA; 439 BP.
XX
AC ACL22483;
XX
DT 27-OCT-2003 (revised)
DT 17-OCT-2003 (first entry)
XX
DE DNA clone originating in barley containing SNP encoding sequence #12474.
XX
KM Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KW gene; ss.
XX
OS Hordeum vulgare; var. (cul.Haruna Nijo).
XX
PN WO2003057877-A1.
XX
PD 17-JUL-2003.
XX
PF 16-DEC-2002; 2002WO-IB005403.
XX
PR 20-DEC-2001; 2001JP-00387059.
XX
PR 20-DEC-2001; 2001JP-00387131.
XX
PR 20-DEC-2001; 2001JP-00403299.
XX
PR 20-DEC-2001; 2001JP-00403300.
XX
PR 27-SEP-2002; 2002JP-00327515.
XX
PA (UYN1-) UNIV JAPAN OKAYAMA.
XX
PI Sato K, Takeda K, Kohara Y;
XX
DR WPI; 2003-587127/55.
XX
PT Single nucleotide polymorphism sites in barley varieties and DNA
PT sequences containing them for analysis and identification of barley
PT varieties and production of barley transformants with desired
PT characteristics.
XX
PS Disclosure; SEQ ID XX; 284bp; Japanese.
XX
CC The present invention relates to oligonucleotide clones originating in
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
CC varieties, identification of particular varieties and genotype-phenotype
CC analysis, isolation of specific genes and creation of new varieties by
CC transformation of barley varieties with them and production of new barley
CC oligonucleotide clone DNA sequence featured in the specification. The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 439 BP; 117 A; 106 C; 81 G; 134 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.: 2.31 Length: 439
Score: 10.00 Matches: 10
```

Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 3.64%  
 DB: 9  
 Gaps: 0  
 Indels: 0  
 Mismatches: 0

US-10-069-353a-8 (1-275) x ACL22483 (1-439)

OY 148 GIUProAspArgAlaIleArgGluIleLeu 157  
 DB 153 GAACCGATCGTCGATTCGGAGATTTTA 124

RESULT 10  
 ID ACL22491/C  
 XX  
 AC ACL22491;  
 XX  
 AC 27-OCT-2003 (revised)  
 DT 17-OCT-2003 (first entry)  
 XX  
 DE DNA clone originating in barley containing SNP encoding sequence #12482.  
 XX  
 KW Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;  
 XX gene; ss.  
 XX  
 OS Hordeum vulgare; var. (cul.Haruna Ni'jo).  
 XX  
 PN WO2003057877-A1.  
 XX  
 PD 17-JUL-2003.  
 XX  
 PF 16-DEC-2002; 2002WO-1B005403.  
 XX  
 PR 20-DEC-2001; 2001JP-00387059.  
 XX  
 PR 20-DEC-2001; 2001JP-00387131.  
 XX  
 PR 20-DEC-2001; 2001JP-00403299.  
 XX  
 PR 20-DEC-2001; 2001JP-00403300.  
 XX  
 PR 27-SEP-2002; 2002JP-00327515.  
 XX  
 PA (UYNI-) UNIV JAPAN OKAYAMA.  
 XX  
 PI Sato K, Takeda K, Kohara Y;  
 XX  
 DR WPI; 2003-587127/55.  
 XX  
 XX

Single nucleotide polymorphism sites in barley varieties and DNA  
 PT sequences containing them for analysis and identification of barley  
 PT varieties and production of barley transformants with desired  
 PT characteristics.  
 PT  
 PS Disclosure; SEQ ID XX; 284pp; Japanese.  
 XX  
 XX

The present invention relates to oligonucleotide clones originating in  
 CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms  
 CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley  
 CC varieties, identification of particular varieties and genotype-phenotype  
 CC analysis, isolation of specific genes and creation of new varieties by  
 CC transformation of barley varieties with them and production of new barley  
 CC varieties with desired properties. The present sequence represents an  
 CC oligonucleotide clone DNA sequence featured in the specification. The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to  
 CC standardise OS field)  
 CC  
 XX

SO Sequence 552 BP; 147 A; 137 C; 102 G; 165 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.: 2.87 Length: 552  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.64% Indels: 0

DB: 9 Gaps: 0

US-10-069-353a-8 (1-275) x ACL22491 (1-552)

OY 148 GIUProAspArgAlaIleArgGluIleLeu 157  
 DB 201 GAACCGATCGTCGATTCGGAGATTTTA 172

RESULT 11  
 ID ACL22487/C  
 XX  
 AC ACL22487;  
 XX  
 AC 27-OCT-2003 (revised)  
 DT 17-OCT-2003 (first entry)  
 XX  
 DE DNA clone originating in barley containing SNP encoding sequence #12478.  
 XX  
 KW Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;  
 XX gene; ss.  
 XX  
 OS Hordeum vulgare; var. (cul.Haruna Ni'jo).  
 XX  
 PN WO2003057877-A1.  
 XX  
 PD 17-JUL-2003.  
 XX  
 PF 16-DEC-2002; 2002WO-1B005403.  
 XX  
 PR 20-DEC-2001; 2001JP-00387059.  
 XX  
 PR 20-DEC-2001; 2001JP-00387131.  
 XX  
 PR 20-DEC-2001; 2001JP-00403299.  
 XX  
 PR 20-DEC-2001; 2001JP-00403300.  
 XX  
 PR 27-SEP-2002; 2002JP-00327515.  
 XX  
 PA (UYNI-) UNIV JAPAN OKAYAMA.  
 XX  
 PI Sato K, Takeda K, Kohara Y;  
 XX  
 DR WPI; 2003-587127/55.  
 XX  
 XX

Single nucleotide polymorphism sites in barley varieties and DNA  
 PT sequences containing them for analysis and identification of barley  
 PT varieties and production of barley transformants with desired  
 PT characteristics.  
 PT  
 PS Disclosure; SEQ ID XX; 284pp; Japanese.  
 XX  
 XX

The present invention relates to oligonucleotide clones originating in  
 CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms  
 CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley  
 CC varieties, identification of particular varieties and genotype-phenotype  
 CC analysis, isolation of specific genes and creation of new varieties by  
 CC transformation of barley varieties with them and production of new barley  
 CC varieties with desired properties. The present sequence represents an  
 CC oligonucleotide clone DNA sequence featured in the specification. The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to  
 CC standardise OS field)  
 CC  
 XX

SO Sequence 572 BP; 151 A; 145 C; 110 G; 166 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2.97 Length: 572  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.64% Indels: 0  
 DB: 9 Gaps: 0

US-10-069-353a-8 (1-275) x ACL22487 (1-572)

```

Db      171  GAACCGGATGTCGATTTCGGGAGATTTTA 142
RESULT 13
ACL22454/c
ID      ACL22454 standard; DNA; 617 BP.
XX
XX
AC      ACL22454;
XX
XX
DT      27-OCT-2003 (revised)
DT      17-OCT-2003 (first entry)
XX
XX
DE      DNA clone originating in barley containing SNP encoding sequence #12445.
XX
XX
KW      Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KW      gene; ss.
XX
XX
OS      Hordeum vulgare; var. (cul.)Haruna Nijo).
XX
XX
PN      WO2003057877-A1.
XX
PD      17-JUL-2003.
XX
PF      16-DEC-2002; 2002WO-IB005403.
XX
PR      20-DEC-2001; 2001JP-00387059.
PR      20-DEC-2001; 2001JP-00387131.
PR      20-DEC-2001; 2001JP-00403299.
PR      20-DEC-2001; 2001JP-00403300.
PR      27-SEP-2002; 2002JP-00327515.
XX
XX
PA      (UTNI-) UNIV JAPAN OKAYAMA.
XX
XX
PI      Sato K, Takeda K, Kohara Y;
XX
XX
DR      WPI; 2003-587127/55.
XX
XX
PT      Single nucleotide polymorphism sites in barley varieties and DNA
PT      sequences containing them for analysis and identification of barley
PT      varieties and production of barley transformants with desired
PT      characteristics.
XX
XX
PS      Disclosure; SEQ ID XX; 284bp; Japanese.
XX
XX
CC      The present invention relates to oligonucleotide clones originating in
CC      barley (Hordeum vulgare) which contain single nucleotide polymorphisms
CC      (SNP). The oligonucleotides may be used for analysis of SNPs among barley
CC      varieties, identification of particular varieties and genotype-phenotype
CC      analysis, isolation of specific genes and creation of new varieties by
CC      transformation of barley varieties with them and production of new barley
CC      varieties with desired properties. The present sequence represents an
CC      oligonucleotide clone DNA sequence featured in the specification. The
CC      sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to
CC      standardise OS field)
XX
XX
SQ      Sequence 617 BP; 160 A; 154 C; 125 G; 178 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      3.19      Length:      617
Score:          10.00     Matches:      10
Percent Similarity: 100.00% Conservative: 0
Best local Similarity: 100.00% Mismatches: 0
Query Match:    3.64%     Indels:      0
DB:            9         Gaps:        0

US-10-069-353A-8 (1-275) x ACL22454 (1-617)
CY      148  GluProAspArgAlaIleLeuGluLeu 157
Db      190  GAACCGGATGTCGATTTCGGGAGATTTTA 161
RESULT 14

```

```

ACL22486/c
ID  ACL22486 standard; DNA: 632 BP.
XX
XX  ACL22486;
AC
XX  27-OCT-2003 (revised)
DT  17-OCT-2003 (first entry)
XX
DE  DNA clone originating in barley containing SNP encoding sequence #12477.
XX
XX  Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KM  gene; ss.
XX
OS  Hordeum vulgare; var. (cul.Haruna Nijo).
XX
PN  WO2003057877-A1.
XX
PD  17-JUL-2003.
XX
PF  16-DEC-2002; 2002WO-IB005403.
XX
PR  20-DEC-2001; 2001JP-00387059.
XX  20-DEC-2001; 2001JP-00387131.
PR  20-DEC-2001; 2001JP-00403299.
XX  20-DEC-2001; 2001JP-00403300.
PR  27-SEP-2002; 2002JP-00327515.
XX
XX  (UYNI-) UNIV JAPAN OKAYAMA.
XX
PI  Sato K, Takeda K, Kohara Y;
DR  WPI; 2003-587127/55.
XX
PT  Single nucleotide polymorphism sites in barley varieties and DNA
PT  sequences containing them for analysis and identification of barley
PT  varieties and production of barley transformants with desired
PT  characteristics.
XX
XX  Disclosure; SEQ ID XX; 284bp; Japanese.
XX
XX  The present invention relates to oligonucleotide clones originating in
XX  barley (Hordeum vulgare) which contain single nucleotide polymorphisms
XX  (SNP). The oligonucleotides may be used for analysis of SNPs among barley
XX  varieties, identification of particular varieties and genotype-phenotype
XX  analysis, isolation of specific genes and creation of new varieties by
XX  transformation of barley varieties with them and production of new barley
XX  varieties with desired properties. The present sequence represents an
XX  oligonucleotide clone DNA sequence featured in the specification. The
XX  sequence data for this patent did not form part of the printed
XX  specification, but was obtained in electronic format directly from WIPO
XX  at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to
XX  standardise OS field)
XX
SQ  Sequence 632 BP; 164 A; 156 C; 130 G; 182 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 3.26 Length: 632
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 9 Gaps: 0
US-10-069-353a-8 (1-275) x ACL22486 (1-632)
QY 148 G|U|P|O|A|S|P|A|R|G|A|A|I|L|E|A|R|G|U|I|L|E|U 157
Db 193 GAACCGATCGTCGATTCGGGAGATTTTA 164
RESULT 15
ACL22486/c
ID  ACL22484 standard; DNA: 633 BP.
XX

```

```

AC  ACL22484;
XX
XX  27-OCT-2003 (revised)
DT  17-OCT-2003 (first entry)
XX
DE  DNA clone originating in barley containing SNP encoding sequence #12475.
XX
XX  Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KM  gene; ss.
XX
OS  Hordeum vulgare; var. (cul.Haruna Nijo).
XX
PN  WO2003057877-A1.
XX
PD  17-JUL-2003.
XX
PF  16-DEC-2002; 2002WO-IB005403.
XX
PR  20-DEC-2001; 2001JP-00387059.
XX  20-DEC-2001; 2001JP-00387131.
PR  20-DEC-2001; 2001JP-00403299.
XX  20-DEC-2001; 2001JP-00403300.
PR  27-SEP-2002; 2002JP-00327515.
XX
XX  (UYNI-) UNIV JAPAN OKAYAMA.
XX
PI  Sato K, Takeda K, Kohara Y;
DR  WPI; 2003-587127/55.
XX
PT  Single nucleotide polymorphism sites in barley varieties and DNA
PT  sequences containing them for analysis and identification of barley
PT  varieties and production of barley transformants with desired
PT  characteristics.
XX
XX  Disclosure; SEQ ID XX; 284bp; Japanese.
XX
XX  The present invention relates to oligonucleotide clones originating in
XX  barley (Hordeum vulgare) which contain single nucleotide polymorphisms
XX  (SNP). The oligonucleotides may be used for analysis of SNPs among barley
XX  varieties, identification of particular varieties and genotype-phenotype
XX  analysis, isolation of specific genes and creation of new varieties by
XX  transformation of barley varieties with them and production of new barley
XX  varieties with desired properties. The present sequence represents an
XX  oligonucleotide clone DNA sequence featured in the specification. The
XX  sequence data for this patent did not form part of the printed
XX  specification, but was obtained in electronic format directly from WIPO
XX  at ftp.wipo.int/pub/published-pct-sequences. (Updated on 27-OCT-2003 to
XX  standardise OS field)
XX
SQ  Sequence 633 BP; 165 A; 156 C; 130 G; 182 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 3.27 Length: 633
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 9 Gaps: 0
US-10-069-353a-8 (1-275) x ACL22484 (1-633)
QY 148 G|U|P|O|A|S|P|A|R|G|A|A|I|L|E|A|R|G|U|I|L|E|U 157
Db 194 GAACCGATCGTCGATTCGGGAGATTTTA 165
Search completed: August 29, 2005, 22:45:39
Job time : 673 secs

```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 29, 2005, 22:18:55 ; Search time 3540 Seconds  
(Without alignments)  
2956.970 Million cell updates/sec

Title: US-10-069-353A-8

Perfect score: 275

Sequence: 1 VLFGAPATSGQVQGMVDLVT.....YEKYADMGYAILTPAKPVG 275

Scoring table:

OLIGO	
Xgapop 60.0 , Xgapext 60.0	
Ygapop 60.0 , Ygapext 60.0	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 34239544 seqs, 19032134700 residues

Word size: 1

Total number of hits satisfying chosen parameters: 68477535

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xjh  
-O=/cgn2.1/USPRO.spool/US10069353/runat.26082005.172152.15063/app\_query.fasta\_1.455  
-DB=EST -OPMT=fastcap -SUPFIX=art -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=plo  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US10069353 @CGN 1.1 3437 @runat.26082005.172152.15063 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOUTERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:\*

1:	gb_est1:*
2:	gb_est2:*
3:	gb_hrc:*
4:	gb_est3:*
5:	gb_est4:*
6:	gb_est5:*
7:	gb_est6:*
8:	gb_g881:*
9:	gb_g882:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	3.6	288	7	CR802983 NF24d11f4
2	10	3.6	324	6	CD662504 UCRV18.0
3	10	3.6	351	1	AV934664 AV934664
4	10	3.6	397	4	B0552337 B0552337
5	10	3.6	406	4	B0548396 B0548396
6	10	3.6	433	4	BG909505 BG909505
7	10	3.6	439	1	AV936770 Talr1104f
8	10	3.6	451	6	CD934354 GR45.123P
9	10	3.6	504	6	CD881537 P1.103118

C 10	10	3.6	511	4	B1750759	B1750759 Ta01.0960
C 11	10	3.6	540	5	B0620788	B0620788 Talr1111C
C 12	10	3.6	552	1	B0552981	B0552981 B0552981
C 13	10	3.6	564	6	CD876873	CD876873 AZ03.111C
C 14	10	3.6	567	1	B0459101	B0459101 B0459101
C 15	10	3.6	572	1	AV945873	AV945873 AV945873
C 16	10	3.6	574	4	B0451604	B0451604 B0451604
C 17	10	3.6	576	6	CD862060	CD862060 AZ01.102D
C 18	10	3.6	594	6	CD896882	CD896882 G174.1048
C 19	10	3.6	597	4	BG904766	BG904766 Talr1134G
C 20	10	3.6	607	4	B0550817	B0550817 B0550817
C 21	10	3.6	617	4	B0546846	B0546846 B0546846
C 22	10	3.6	632	1	AV944357	AV944357 AV944357
C 23	10	3.6	633	1	AV942357	AV942357 AV942357
C 24	10	3.6	634	1	B0548698	B0548698 B0548698
C 25	10	3.6	637	6	CA597128	CA597128 wpa1c.pk0
C 26	10	3.6	642	4	B0548280	B0548280 B0548280
C 27	10	3.6	651	4	B0547951	B0547951 B0547951
C 28	10	3.6	653	6	CD934353	CD934353 GR45.123P
C 29	10	3.6	674	4	B0552252	B0552252 B0552252
C 30	10	3.6	685	4	B0548223	B0548223 B0548223
C 31	10	3.6	700	1	AV936690	AV936690 AV936690
C 32	10	3.6	703	4	B0547277	B0547277 B0547277
C 33	10	3.6	705	4	B0548319	B0548319 B0548319
C 34	10	3.6	705	4	B0552839	B0552839 B0552839
C 35	10	3.6	708	1	AV942972	AV942972 AV942972
C 36	10	3.6	719	4	B0546326	B0546326 B0546326
C 37	10	3.6	721	4	B0548517	B0548517 B0548517
C 38	10	3.6	726	4	B0548683	B0548683 B0548683
C 39	10	3.6	734	9	CE354658	CE354658 l1gr-g88-
C 40	10	3.6	821	2	BF065038	BF065038 HV_CBD002
C 41	10	3.6	951	2	BF255786	BF255786 B05MEF00
C 42	9	3.3	139	1	A1061356	A1061356 a008A03.8
C 43	9	3.3	236	7	CF681341	CF681341 CCA879TR
C 44	9	3.3	375	1	A0312420	A0312420 A0312420
C 45	9	3.3	397	6	CA735338	CA735338 wpl1b.pk0

#### ALIGNMENTS

RESULT 1  
CR802983/c  
LOCUS  
DEFINITION  
NP24d11f4.r1 Tall Fescue P1297901 44 deg C Heat Stress SSH cDNA  
Schedonorus arundinaceus cDNA clone NP24d11f4 5', mRNA sequence.  
CR802983  
VERSION  
ACCESSION  
CR802983.1 GI:43404841  
KEYWORDS  
SOURCE  
ORGANISM

Schedonorus arundinaceus (Festuca arundinacea)  
EST.  
Schedonorus arundinaceus  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Poae; Schedonorus.  
1 (bases 1 to 288)

REFERENCE  
Zhang, Y., Zwonitzer, J. C., Chekhovskiy, K., May, G. D. and Mian, M. A. R.  
A functional genomics approach for identification of heat tolerance  
genes in tall fescue  
(in) Hopkin, A., Wang, Z. Y., Mian, R., Sledge, M. and Barker, R. E.  
(eds.), MOLECULAR BREEDING OF FORAGE AND TURF, Kluwer Acad. Pub. 0,  
87-96 (2003)

COMMENT  
Contact: Bruce A. Roe, University of Oklahoma, broeou@ou.edu  
Department of Chemistry and Biochemistry  
Advanced Center for Genome Technology, University of Oklahoma  
620 Parrington Oval, Norman, OK 73019, USA  
Tel: 405 325 4912  
Fax: 405 325 7762

FEATURES  
source  
Contact: Dr. Rouf Mian (rmian@noble.org) regarding clone availability  
Seq primer: M13 reverse primer  
High quality sequence stop: 281.  
Location/Qualifiers  
1..288  
/organism="Schedonorus arundinaceus"

```

/mo1_type="mRNA"
/cultivar="P1297901"
/db_xref="taxon:4606"
/clone="NF24d11f44"
/issue_type="shoot"
/lab_host="E.coli"
/clone_1ib="Tail Rescue P1297901 44 deg C Heat Stresses SSH
cDNA"
/note="Vector: PCR 2.1-TOPO; Site 1: EcoR I; Site 2: EcoR
I; BD/Clontech PCR-select cDNA subtraction library"

ORIGIN

Alignment Scores:
Pred. No.:      14.1      Length:      288
Score:          10.00     Matches:      10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:     3.64%     Indels:      0
DB:              7        Gaps:          0

US-10-069-353A-8 (1-275) x CK802983 (1-288)

QY      148 GUProASPARGAla1leArgGluIleu 157
      |||||
Db      270 GAACCGAGCGCGCTATTGCGAGATATTG 241

RESULT 2
LOCUS   CD662504               324 bp      mRNA      linear      EST 23-JUN-2003
DEFINITION   UCRHV18_02af09 b1 Drought-stressed Dicktoo barley epidermis cDNA
              library Hordeum vulgare subsp. vulgare cDNA clone UCRHV18_02af09,
              mRNA sequence.
ACCESSION   CD662504
VERSION     CD662504.1   GI:32152807
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
ORGANISM    Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Poideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 324)
            Wahid,A., Close,T.J., Fenton,R.D., Manamaker,S., Collura,K.,
            Feuerbacher,O., Kim,H.R., Kudrna,D., Wing,R. and Yu,Y.
            Drought-stressed barley leaf epidermis cDNA sequences
            Unpublished (2003)
            Contact: Timothy J. Close
            Department of Botany & Plant Sciences
            University of California
            Riverside, CA 92521-0124, USA
            Tel: 909-787-3318
            Fax: 909-787-4437
            Email: timothy.close@ucr.edu
            Seg primer: T7
FEATURES             Location/Qualifiers
     source          1..324
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Dicktoo"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="UCRHV18_02af09"
                     /tissue_type="lower leaf epidermis"
                     /dev_stage="1-2 week seedlings"
                     /lab_host="E. coli T10121"
                     /clone_1ib="Drought-stressed Dicktoo barley epidermis cDNA
                     library"
                     /note="Vector: lambda Uni-ZAP XR, excised phagemid;
                     Site 1: EcoRI; Site 2: XhoI; Seeds of barley (Hordeum
                     vulgare L. cv. Dicktoo) were germinated in dishes
                     containing UCR-mix soil. Seedlings were kept in a growth
                     chamber at 20C (day/night) and allowed to grow at 60-70%
                     soil moisture content. After 3-4 days, the water was
                     withheld in order to apply drought until the soil moisture

```

content was reduced to 10-12%, which took another 3-4 days. At the time of extraction of epidermis, leaf water and osmotic potentials had dropped to 15.9%, 20.2% and 24.7% of controls, respectively, as determined using a vapor pressure osmometer (Model 5100C, Wescor, Inc., Logan, UT). Epidermal strips were quickly peeled off of seedlings and immediately submerged in liquid nitrogen. About 15-20 g of epidermal tissue was collected and used to extract total RNA. Total RNA was extracted using Concert Plant RNA Reagent (Invitrogen 12322-012). Poly(A) RNA was purified using PolyAtrack mRNA Isolation System IV (Promega). A cDNA library was made using a Uni-ZAP cDNA synthesis kit (Stratagene). A total of 1 million primary lambda cDNA clones were mass-excised in vitro to give a population of phagescript SK(-) phagemids. All steps to this point were performed in the T7 Close lab at the University of California, Riverside by A. Wahid with some assistance from R.D. Fenton. Phagemids were plated on the T7C121 host strain, plasmid DNA purified, cDNA clones archived, and DNA sequences determined using the T7 primer (mainly 3' end reads) using an ABI3730 at the Arizona Genomics Institute, University of Arizona (Collura, Feuerbacher, Kim, Kudrna, Wing, Yu). Chromatogram files were transmitted to UC Riverside (by Yu), then processed at UC Riverside (by Manamaker) using the HarVEST pipeline (http://harvest.ucr.edu) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to Genbank."

```

ORIGIN

Alignment Scores:
Pred. No.:      15.8      Length:      324
Score:          10.00     Matches:      10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:     3.64%     Indels:      0
DB:              6        Gaps:          0

US-10-069-353A-8 (1-275) x CD662504 (1-324)

QY      148 GUProASPARGAla1leArgGluIleu 157
      |||||
Db      167 GAACCGAGTCGCGATTGCGAGATTTTA 138

RESULT 3
LOCUS     AV934664               351 bp      mRNA      linear      EST 18-JAN-2002
DEFINITION   AV934664 K. Sato unpublished cDNA library, cv. Haruna Nijo adult,
              heading stage top three leaves Hordeum vulgare subsp. vulgare cDNA
              clone baal1d01 3', mRNA sequence.
ACCESSION   AV934664
VERSION     AV934664.1   GI:18230461
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
ORGANISM    Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Poideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 351)
            Sato,K., Saitoh,D. and Takeda,K.
            Barley EST sequencing project in NIG and Okayama Univ
            Unpublished (2002)
            Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshin@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..351

```



/organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cultivar="Haruna NiJo"  
 /sub\_species="vulgare"  
 /db\_xref="taxon:112509"  
 /clone="baal1d01"  
 /tissue\_type="top three leaves"  
 /dev\_stage="adult, heading stage"  
 /clone\_1ib="K. Sato unpublished cDNA library, cv. Haruna NiJo adult, heading stage top three leaves"

## ORIGIN

## Alignment Scores:

Pred. No.:	17	Length:	351
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.64%	Indels:	0
DB:	1	Gaps:	0

US-10-069-353a-8 (1-275) x AV934664 (1-351)

QY 148 GluProAspArgAlaIleArgGluIleLeu 157  
 |||||  
 169 GAACCGGATCGTCGATTCCGGAGATTTTA 140

RESULT 4 397 bp mRNA linear EST 14-NOV-2002  
 BU552337/c  
 LOCUS BU552337 K. Sato unpublished cDNA library, strain H602 adult,  
 DEFINITION heading stage top three leaves Hordeum vulgare subsp. spontaneum  
 CDS  
 BU552337  
 BU552337.1 GI:24970788  
 EST.  
 Hordeum vulgare subsp. spontaneum  
 Hordeum vulgare subsp. spontaneum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Hordeum.  
 1 (bases 1 to 397)  
 Sato, K., Saitoh, D. and Takeda, K.  
 Barley EST sequencing project in NIG and Okayama Univ  
 Unpublished (2002)  
 Contact: Tadao Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshin@genes.nig.ac.jp.  
 Location/Qualifiers

FEATURES  
 source  
 1..397  
 /organism="Hordeum vulgare subsp. spontaneum"  
 /mol\_type="mRNA"  
 /strain="H602"  
 /sub\_species="spontaneum"  
 /db\_xref="taxon:77009"  
 /clone="bah52m15"  
 /tissue\_type="top three leaves"  
 /dev\_stage="adult, heading stage"  
 /clone\_1ib="K. Sato unpublished cDNA library, strain H602 adult, heading stage top three leaves"

REFERENCE  
 AUTHORS Sato, K., Saitoh, D. and Takeda, K.  
 TITLE Barley EST sequencing project in NIG and Okayama Univ  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Tadao Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshin@genes.nig.ac.jp.  
 Location/Qualifiers

FEATURES  
 source  
 1..397  
 /organism="Hordeum vulgare subsp. spontaneum"  
 /mol\_type="mRNA"  
 /strain="H602"  
 /sub\_species="spontaneum"  
 /db\_xref="taxon:77009"  
 /clone="bah52m15"  
 /tissue\_type="top three leaves"  
 /dev\_stage="adult, heading stage"  
 /clone\_1ib="K. Sato unpublished cDNA library, strain H602 adult, heading stage top three leaves"

## ORIGIN

## Alignment Scores:

Pred. No.:	19	Length:	397
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.64%	Indels:	0
DB:	4	Gaps:	0

US-10-069-353a-8 (1-275) x BU552337 (1-397)

QY 148 GluProAspArgAlaIleArgGluIleLeu 157  
 |||||  
 DB 238 GAACCGGATCGTCGATTCCGGAGATTTTA 209

RESULT 5 406 bp mRNA linear EST 14-NOV-2002  
 BU548396/c  
 LOCUS BU548396 K. Sato unpublished cDNA library, cv. Haruna NiJo adult,  
 DEFINITION heading stage top three leaves Hordeum vulgare subsp. vulgare cDNA  
 clone baal24d16 3', mRNA sequence.

ACCESSION BU548396  
 VERSION BU548396.1 GI:24966834  
 KEYWORDS EST.  
 SOURCE Hordeum vulgare subsp. vulgare  
 ORGANISM Hordeum vulgare subsp. vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Hordeum.  
 1 (bases 1 to 406)  
 Sato, K., Saitoh, D. and Takeda, K.  
 Barley EST sequencing project in NIG and Okayama Univ  
 Unpublished (2002)  
 Contact: Tadao Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshin@genes.nig.ac.jp.  
 Location/Qualifiers

## FEATURES

source  
 1..406  
 /organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cultivar="Haruna NiJo"  
 /sub\_species="vulgare"  
 /db\_xref="taxon:112509"  
 /clone="baal24d16"  
 /tissue\_type="top three leaves"  
 /dev\_stage="adult, heading stage"  
 /clone\_1ib="K. Sato unpublished cDNA library, cv. Haruna NiJo adult, heading stage top three leaves"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 19.4  
 Score: 10.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 3.64%  
 DB: 4  
 Length: 406  
 Matches: 10  
 Conservative: 0  
 Mismatches: 0  
 Indels: 0  
 Gaps: 0

US-10-069-353a-8 (1-275) x BU548396 (1-406)

QY 148 GluProAspArgAlaIleArgGluIleLeu 157  
 |||||  
 DB 168 GAACCGGATCGTCGATTCCGGAGATTTTA 139

RESULT 6 433 bp mRNA linear EST 05-JUN-2001  
 BG909505/c  
 LOCUS BG909505 Triticum aestivum cDNA clone TAlr1104F08 5',  
 DEFINITION TAlr1104F08 TAlr1 Triticum aestivum cDNA clone TAlr1104F08 5',  
 mRNA sequence.  
 ACCESSION BG909505  
 VERSION BG909505.1 GI:14317181  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 1 (bases 1 to 433)

## REFERENCE

AUTHORS Cloutier, S., Dong, G. and Walsh, A.  
 TITLE Wheat functional genomics - Thatcher Lr1 cDNA library  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Dr. Sylvie Cloutier  
 Cereal Research Centre, Agriculture and Agri-food Canada  
 195 Dufresne Rd, Winnipeg, MB, Canada R3T 2M9  
 Tel: (204) 983-2340  
 Fax: (204) 983-4604  
 Email: scloutier@gr.gc.ca

was cloned directionally, not all sequences generated with reverse primer were from the 5' end (same with forward primer and 3' end).  
 Average insert size is >2.2 kb  
 Plate: 104 row: F column: 08  
 Seq primer: M13 Reverse.

FEATURES  
 source Location/Qualifiers

1..433  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Thatcher Lr1"  
 /db\_xref="taxon:4565"  
 /clone="TALr1104F08"  
 /tissue\_type="leaf tissue"  
 /dev\_stage="14 Days old"  
 /lab\_host="E. coli XL0LR"  
 /clone\_11b="TALr1"  
 /note="Vector: Lambda ZapII; mass excised in plasmid vector PBK-CMV (Stratagene). Site 1: EcoRI; Site 2: XhoI; mRNA obtained from wheat NIL Thatcher Lr1 24 hours after inoculation with leaf rust pathogen Puccinia triticina race BBB carrying the avirulence gene Avr1."

## ORIGIN

## Alignment Scores:

Pred. No.:	20.6	Length:	433
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.64%	Indels:	0
DB:	4	Gaps:	0

US-10-069-353A-8 (1-275) x BG909505 (1-433)

Qy 148 GIUProaspAga1a1leargluleu 157  
 |||||  
 Db 237 GAACCGAGTCGTCGATCGGAGATTGTG 208

## RESULT 7

LOCUS AV936770 439 bp mRNA linear EST 18-JAN-2002  
 AV936770 K. Sato unpublished cDNA library, cv. Haruna Nijo adult,  
 heading stage top three leaves Hordeum vulgare subsp. vulgare cDNA  
 clone baal509 3', mRNA sequence.

ACCESSION AV936770 GI:18232567  
 VERSION AV936770.1  
 KEYWORDS EST.  
 SOURCE Hordeum vulgare subsp. vulgare  
 ORGANISM Hordeum vulgare subsp. vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Hordeum.

REFERENCE 1 (bases 1 to 439)

AUTHORS Sato, K., Saitoh, D. and Takeda, K.  
 TITLE Barley EST sequencing project in NIG and Okayama Univ  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Tadao Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshini@gens.nig.ac.jp.

FEATURES  
 source Location/Qualifiers  
 1..439

/organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cultivar="Haruna Nijo"  
 /sub\_species="vulgare"  
 /db\_xref="taxon:112509"  
 /clone="baal509"  
 /tissue\_type="top three leaves"  
 /dev\_stage="adult, heading stage"  
 /clone\_11b="K. Sato unpublished cDNA library, cv. Haruna Nijo adult, heading stage top three leaves"

## ORIGIN

## Alignment Scores:

Pred. No.:	20.8	Length:	439
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.64%	Indels:	0
DB:	1	Gaps:	0

US-10-069-353A-8 (1-275) x AV936770 (1-439)

Qy 148 GIUProaspAga1a1leargluleu 157  
 |||||  
 Db 153 GAACCGAGTCGTCGATCGGAGATTGTA 124

## RESULT 8

LOCUS CD934354 451 bp mRNA linear EST 15-JUL-2003  
 CD934354 GR45.123P11R010830 GR45 Triticum aestivum cDNA clone GR45123P11,  
 mRNA sequence.

ACCESSION CD934354  
 VERSION CD934354.1 GI:32782118  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 451)

AUTHORS Genoplante, a major partnership french program in plant genomics  
 TITLE Unpublished (2003)  
 JOURNAL Contact: Genoplante  
 COMMENT Genoplante  
 93, rue Henri Rochefort 91025 EVRY CEDEX France  
 Tel: 33 1 69 47 54 00  
 Fax: 33 1 69 47 54 10  
 This sequence has been generated in the framework of the french  
 plant genomics programme 'Genoplante' (<http://www.genoplante.com>  
 and <http://genoplante-info.infobiogen.fr>).

FEATURES  
 source Location/Qualifiers  
 1..451  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="recital"  
 /db\_xref="taxon:4565"  
 /clone="GR45123P11"  
 /tissue\_type="grain (45 degrees per day after  
 pollination)"  
 /clone\_11b="GR45"

## ORIGIN

## Alignment Scores:

Pred. No.:	21.4	Length:	451
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.64%	Indels:	0
DB:	6	Gaps:	0

US-10-069-353A-8 (1-275) x CD934354 (1-451)

QY 154 ArgGlut1leuArgValleuLysProGly 163  
 |||||  
 Db 388 CGAGAGATCCTTAGAGTCTCTAAACACGAGA 359

RESULT 9  
 LOCUS CD881537 504 bp mRNA linear EST 14-JUL-2003  
 DEFINITION F1.103118P010329 F1 Triticum aestivum cDNA clone F1103118, mRNA  
 sequence.  
 ACCESSION CD881537  
 VERSION CD881537.1 GI:32641050  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 1 (bases 1 to 504)  
 REFERENCE  
 AUTHORS Genoplante.  
 TITLE Genoplante, a major partnership french program in plant genomics  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Genoplante  
 Genoplante  
 93, rue Henri Rochefort 91025 EVRY CEDEX France  
 Tel: 33 1 69 47 54 00  
 Fax: 33 1 69 47 54 10  
 This sequence has been generated in the framework of the french  
 plant genomics programme 'Genoplante' (<http://www.genoplante.com>  
 and <http://genoplante-info.infobiogen.fr>).

FEATURES  
 source  
 1..504  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="rectal"  
 /db\_xref="taxon:4565"  
 /clone="F1103118"  
 /tissue\_type="leaf one"  
 /clone\_lib="F1"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 23 7 Length: 504  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.64% Indels: 0  
 DB: 6 Gaps: 0

US-10-069-353a-8 (1-275) x CD881537 (1-504)

QY 148 GluProAspArgAlaIleArgGluIleLeu 157  
 |||||  
 Db 267 GAGCCAGACCGTGTATTCCGGAGATTG 296

RESULT 10  
 LOCUS BI750759/c 511 bp mRNA linear EST 25-SEP-2001  
 DEFINITION Ta01\_09G01.R  
 Ta01\_AAFc\_EOORC\_Fusarium graminearum inoculated wheat heads  
 Triticum aestivum cDNA clone Ta01\_09G01, mRNA sequence.  
 ACCESSION BI750759  
 VERSION BI750759.1 GI:15772561  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 1 (bases 1 to 511)  
 REFERENCE  
 AUTHORS Ouellet, T., Dan, H., Koul, A., Chapados, V., Couroux, P., De Moors, A.,  
 Harris, L. J., Hattori, J. I., Robert, L. S., Singh, J. A., Sprout, D. and  
 Tinker, N. A.  
 TITLE Expressed Sequence Tags from Wheat Heads 24 Hours after Spray

JOURNAL  
 COMMENT Unpublished (2001)  
 Contact: Ouellet, Therese  
 Eastern Cereal and Oilseed Research Centre  
 Agriculture and Agri-food Canada  
 Neaby Bldg., Central Experimental Farm, Ottawa, Ontario, KIA 0C6,  
 CANADA  
 Tel: (613) 759-1658  
 Fax: (613) 759-1701  
 Email: [ouellet@agr.gc.ca](mailto:ouellet@agr.gc.ca).

FEATURES  
 source  
 1..511  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Frontana"  
 /db\_xref="taxon:4565"  
 /clone="Ta01\_09G01"  
 /tissue\_type="heads"  
 /dev\_stage="anthesis"  
 /clone\_lib="Ta01\_AAFc\_EOORC\_Fusarium graminearum inoculate  
 d\_wheat\_heads"  
 /note="Vector: pGEM-T easy; Site\_1: EcoRI; Site\_2: EcoRI;  
 Controlled chamber-grown wheat heads were spray inoculated  
 at mid-anthesis with a Fusarium graminearum macroconidial  
 suspension (50,000 spores/ml) and kept under intermittent  
 misting for 24 hours, then collected and immediately  
 frozen in liquid nitrogen."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 24 Length: 511  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.64% Indels: 0  
 DB: 4 Gaps: 0

US-10-069-353a-8 (1-275) x BI750759 (1-511)

QY 148 GluProAspArgAlaIleArgGluIleLeu 157  
 |||||  
 Db 159 GAACCGACCGTCGATTCGGAGATTG 130

RESULT 11  
 LOCUS BQ620788/c 540 bp mRNA linear EST 28-JUN-2002  
 DEFINITION Talr1111C01R Talr1 Triticum aestivum cDNA clone Talr1111C01R, mRNA  
 sequence.  
 ACCESSION BQ620788  
 VERSION BQ620788.1 GI:21625867  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 1 (bases 1 to 540)  
 REFERENCE  
 AUTHORS Cloutier, S., Dong, G. and Walsh, A.  
 TITLE Wheat functional genomics - Thatcher Lr1 cDNA library  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Dr. Sylvie Cloutier  
 Cereal Research Centre, Agriculture and Agri-food Canada  
 195 Datoe Rd, Winnipeg, MB, Canada R3T 2N9  
 Tel: (204) 983-2340  
 Fax: (204) 983-4604  
 Email: [scloutier@agr.gc.ca](mailto:scloutier@agr.gc.ca)  
 was cloned directionally, not all sequences generated with reverse  
 primer where from the 5' end (same with forward primer and 3' end).  
 Average insert size is >2.2kb  
 Plate: 111 row: C column: 01  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1..540

/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Thatcher Lr1"  
/db\_xref="taxon:4555"  
/clone="TAB111C01R"  
/tissue\_type="Leaf tissue"  
/dev\_stage="14 Days old"  
/lab\_host="E. coli XL0LR"  
/clone\_lib="Tair1"  
/note="Vector: lambda ZapII; mass excised in plasmid  
vector pK-CMV (Stratagene). Site\_1: EcoRI; Site\_2: XhoI;  
mRNA obtained from wheat NIL Thatcher Lr1 24 hours after  
inoculation with leaf rust pathogen Puccinia triticina  
race BBB carrying the avirulence gene Avr1."

## ORIGIN

## Alignment Scores:

Pred. No.:	25.2	Length:	540
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.64%	Indels:	0
DB:	5	Gaps:	0

US-10-069-353A-8 (1-275) x B0620788 (1-540)

Qy 148 GtProAspArgAlaIleArgGluIleLeu 157

Db 230 GAGCCGACCGTCGATTCGGAGATTG 201

## RESULT 12

BU552981 552 bp mRNA linear EST 14-NOV-2002  
LOCUS BU552981 K. Sato unpublished cDNA library, strain H602 adult.  
DEFINITION heading stage top three leaves Hordeum vulgare subsp. spontaneum  
CDNA clone bahe2h02 3', mRNA sequence.

## ACCESSION

BU552981 GI:24971432

## VERSION

BU552981

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Sato, K., Saitoh, D. and Takeda, K.  
Barley EST sequencing project in NIG and Okayama Univ  
Unpublished (2002)  
Contact: Tadao Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers

## FEATURES

## source

1. 552  
/organism="Hordeum vulgare subsp. spontaneum"  
/mol\_type="mRNA"  
/strain="H602"  
/sub\_species="spontaneum"  
/db\_xref="taxon:77009"  
/clone="bahe2h02"  
/tissue\_type="top three leaves"  
/dev\_stage="adult, heading stage"  
/clone\_lib="K. Sato unpublished cDNA library, strain H602  
adult, heading stage top three leaves"

## ORIGIN

## Alignment Scores:

Pred. No.:	25.7	Length:	552
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0

Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.64% Indels: 0  
DB: 4 Gaps: 0

US-10-069-353A-8 (1-275) x BU552981 (1-552)

Qy 148 GtProAspArgAlaIleArgGluIleLeu 157

Db 201 GAGCCGACCGTCGATTCGGAGATTG 172

## RESULT 13

CD876873 564 bp mRNA linear EST 11-JUL-2003  
LOCUS CD876873 C  
DEFINITION AZ03.111C19R011127 AZ03 Triticum aestivum cDNA clone AZ03111C19,  
mRNA sequence.

## ACCESSION

CD876873 GI:32560689

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Triticum aestivum (bread wheat)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
1 (bases 1 to 564)  
Genoplane.  
Unpublished (2003)  
Contact: Genoplane  
Genoplane  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french  
plant genomics programme 'Genoplane' (<http://www.genoplane.com>  
and <http://genoplane-info.infobiogen.fr>).  
Location/Qualifiers

## FEATURES

## source

1. 564  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="rectal"  
/db\_xref="taxon:4565"  
/clone="AZ03111C19"  
/tissue\_type="leaf"  
/clone\_lib="AZ03"

## ORIGIN

## Alignment Scores:

Pred. No.:	26.2	Length:	564
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.64%	Indels:	0
DB:	6	Gaps:	0

US-10-069-353A-8 (1-275) x CD876873 (1-564)

Qy 148 GtProAspArgAlaIleArgGluIleLeu 157

Db 229 GAGCCGACCGTCGATTCGGAGATTG 200

## RESULT 14

BU459101 567 bp mRNA linear EST 23-MAY-2002  
LOCUS BU459101 K. Sato unpublished cDNA library, cv. Akashinriki  
DEFINITION vegetative stage leaves Hordeum vulgare subsp. vulgare cDNA clone  
baak34i03 3', mRNA sequence.

## ACCESSION

BU459101 GI:21137635

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

REFERENCE Poideae; Triticaceae; Hordeum.  
1 (bases 1 to 567)  
AUTHORS Sato, K., Saisho, D. and Takeda, K.  
TITLE Barley EST sequencing project in NIG and Okayama Univ  
JOURNAL Unpublished (2002)  
COMMENT Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp.

FEATURES  
source  
1..567  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Akashinriki"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="baak34103"  
/issue\_type="leaves"  
/dev\_stage="vegetative stage"  
/clone\_lib="K. Sato unpublished cDNA library, cv.  
Akashinriki vegetative stage leaves"

## ORIGIN

## Alignment Scores:

Pred. No.:	26.4	Length:	567
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.64%	Indels:	0
DB:	4	Gaps:	0

US-10-069-353a-8 (1-275) x BJ459101 (1-567)

QY 154 ArgGluIleLeuArgValLeuLysProGly 163  
|||||  
Db 372 CGAGAGATCTGAGAGTCTAAACACGCG 343

RESULT 15 572 bp mRNA linear EST 18-JAN-2002  
AV945873/c LOCUS  
AV945873 K. Sato unpublished cDNA library, strain H602 adult,  
DEFINITION heading stage top three leaves Hordeum vulgare subsp. spontaneum  
cDNA clone bah28k06 3', mRNA sequence.

ACCESSION AV945873  
VERSION AV945873.1 GI:18241670  
KEYWORDS EST.  
SOURCE Hordeum vulgare subsp. spontaneum  
ORGANISM Hordeum vulgare subsp. spontaneum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticaceae; Hordeum.  
1 (bases 1 to 572)

REFERENCE Sato, K., Saisho, D. and Takeda, K.  
TITLE Barley EST sequencing project in NIG and Okayama Univ  
JOURNAL Unpublished (2002)  
COMMENT Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp.

FEATURES  
source  
1..572  
/organism="Hordeum vulgare subsp. spontaneum"  
/mol\_type="mRNA"  
/strain="H602"  
/sub\_species="spontaneum"  
/db\_xref="taxon:77009"  
/clone="bah28k06"

## ORIGIN

/tissue\_type="top three leaves"  
/dev\_stage="adult, heading stage"  
/clone\_lib="K. Sato unpublished cDNA library, strain H602  
adult, heading stage top three leaves"

## Alignment Scores:

Pred. No.:	26.6	Length:	572
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.64%	Indels:	0
DB:	1	Gaps:	0

US-10-069-353a-8 (1-275) x AV945873 (1-572)

QY 148 GIUProASPARGAlaIleArgGluIleLeu 157  
|||||  
Db 201 GAACCGATCGTGCATTCGGGAGATTTTA 172

Search completed: August 30, 2005, 01:03:29  
Job time : 3551 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 29, 2005, 22:25:40 ; Search time 227 Seconds  
(without alignments)  
1982.274 Million cell updates/sec

Title: US-10-069-353A-8

Perfect score: 275 1 VLPGAPTSQGVGMVDTVT.....YKVAHDMGXALITARKPVG 275

## Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2400006

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

## Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O/cgn2\_1/USPTO.spool/US10069353/runat.26082005.172152.15080/app.query.fasta\_1.455  
-DB=Issued Patents.NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOFCG=0  
-LOOPEXT=0 -UNITs=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10069353 @CEN 1.1 69 @runat.26082005.172152.15080 -NCPU=6 -ICPU=3  
-NO MAP -LARGEOUTPUT -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

## Database :

Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCPUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	275	100.0	80161	3	US-09-036-987A-1
2	275	100.0	80161	3	US-09-370-700-1
3	275	100.0	80161	4	US-09-603-207-1
4	10	3.6	738	4	US-09-724-797-13
5	10	3.6	6085	3	US-09-029-603-4
6	9	3.3	400	4	US-08-956-171E-2979
7	9	3.3	400	4	US-08-781-986A-2979
8	9	3.3	11466	4	US-08-956-171E-444
9	9	3.3	11466	4	US-08-781-986A-444
10	8	2.9	310	4	US-09-270-767-28714
11	8	2.9	414	4	US-09-252-991A-334
12	8	2.9	456	4	US-09-248-796A-3885

C	13	8	2.9	505	4	US-09-621-976-18024	Sequence 18024, A
	14	8	2.9	507	4	US-09-902-540-7484	Sequence 7484, Ap
	15	8	2.9	522	4	US-09-902-540-3761	Sequence 3761, Ap
	16	8	2.9	579	3	US-08-991-789A-147	Sequence 147, App
	17	8	2.9	579	3	US-09-062-451-147	Sequence 147, App
	18	8	2.9	579	3	US-09-598-326-147	Sequence 147, App
	19	8	2.9	579	4	US-09-289-198-147	Sequence 147, App
	20	8	2.9	579	4	US-09-429-755-147	Sequence 147, App
	21	8	2.9	579	4	US-09-699-295-147	Sequence 16384, A
	22	8	2.9	597	4	US-09-252-991A-16384	Sequence 16006, A
	23	8	2.9	630	4	US-09-252-991A-16006	Sequence 7735, Ap
	24	8	2.9	651	4	US-09-252-991A-7735	Sequence 16273, A
	25	8	2.9	660	4	US-09-252-991A-16273	Sequence 637, App
	26	8	2.9	759	4	US-09-583-110-637	Sequence 5226, Ap
	27	8	2.9	768	4	US-09-902-540-5226	Sequence 1646, Ap
	28	8	2.9	780	4	US-09-107-433-1646	Sequence 11034, A
	29	8	2.9	789	4	US-09-252-991A-11034	Sequence 2, Appl1
	30	8	2.9	891	1	US-08-687-895-2	Sequence 2, Appl1
	31	8	2.9	891	2	US-09-040-482-2	Sequence 6460, Ap
	32	8	2.9	933	4	US-09-248-796A-6460	Sequence 2638, Ap
	33	8	2.9	1101	4	US-09-489-039A-2638	Sequence 312, App
	34	8	2.9	1110	4	US-09-252-991A-312	Sequence 7650, Ap
	35	8	2.9	1161	4	US-09-252-991A-7650	Sequence 12864, A
	36	8	2.9	1161	4	US-09-270-767-12864	Sequence 5791, Ap
	37	8	2.9	1212	4	US-09-489-039A-5791	Sequence 13, Appl1
	38	8	2.9	1232	3	US-08-411-760-13	Sequence 1, Appl1
	39	8	2.9	1233	3	US-08-244-491A-1	Sequence 197, App
	40	8	2.9	1233	2	US-08-797-727A-1	Sequence 67, Appl
	41	8	2.9	1282	4	US-09-205-258-197	Sequence 9997, Ap
	42	8	2.9	1401	4	US-09-602-787A-67	Sequence 4219, Ap
	43	8	2.9	1424	4	US-09-270-767-9997	Sequence 353, App
	44	8	2.9	1692	4	US-09-902-540-4219	
	45	8	2.9	1719	4	US-09-252-991A-353	

## ALIGNMENTS

RESULT 1  
US-09-036-987A-1  
; Sequence 1, Application US/09036987A  
; Patent No. 6143526  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H.  
; APPLICANT: Crawford, Mary C.  
; APPLICANT: Crawford, Kathryn P.  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Madduri, Donald J.  
; APPLICANT: Treadway, Patli J.  
; APPLICANT: Turner, Jan R.  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; TITLE OF INVENTION: Production  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dow Agrosciences LLC Patent Department  
; STREET: 9330 Zionsville Road  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: USA  
; ZIP: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/036,987A  
; FILING DATE: 09-MAR-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stuart, Donald R  
; REGISTRATION NUMBER: 28,479  
; REFERENCE/DOCKET NUMBER: 50,608

```
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (317)337-4816
/ TELEFAX: (317)337-4847
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 80161 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

Alignment Scores:
Pred. No.: 5,22e-273 Length: 80161
Score: 275.00 Matches: 275
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-069-353A-8 (1-275) x US-09-036-987A-1 (1-80161)

QY 1 ValLeuProGlyValAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20
DB 20168 GTGTTGCCAGGTGGCGCACCAATCGCAGGTTGGGCAATGATGATGACCTGTGTCACG 20227
QY 21 ProLeuLeuAsnSerValAlaGlyGlyProCysAlaIleHisGlyTyrTrpGluAsn 40
DB 20228 CCGTTCCTGAATCTCGGTCCGGCCGGCCCTTCGCGCATTCACACGGCTACTGGAGAAC 20287
QY 41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
DB 20288 GACGGCGCGGCTTCTCGTGGCAGAGCGCGCCGACCGGCTCACCGACTTGTCCGCAACGG 20347
QY 61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80
DB 20348 ACCGTGCTCGATGGCGGCTTGCACCTGCTGATGTGGGGTGGGTAACCGAACACGCGG 20407
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
DB 20408 CTGGCGCTGCGCGCGACCAACCGCATCCGATCACCGGCTACACCGTCAACCGAGTGCNA 20467
QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
DB 20468 GTGGCCATGCGCGCATGATGCCGACGCGGACCTPAAGCCACCGGGTGGACTTCTCG 20527
QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
DB 20528 TGCCTCGATGCAATGCTCTGCGCGTACCGGACCAATGCTTTCGACGCGCTGGGCGCATG 20587
QY 141 GlnSerLeuLeuGlnMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160
DB 20588 CAGTGGCTGTGGAGATGTCGAACCGGACCGTGCATTCGGGAAATCTTTCGATGACTC 20647
QY 161 LysProGlyGlyIleLeuGlyValThrGluValValLysArgGluAlaGlyIleGlyMet 180
DB 20648 AAACCCGGGTCATCTCGGCGTCAACCGAGTCTCAAAACGAGACGCGGCGCGGATG 20707
QY 181 ProValSerGlyAspArgTrpProThrArgIleLeuArgIleCysLeuAlaGluGlnLeu 200
DB 20708 CCGGGTCCGGGACAGGTGGCGGACCGGCTTCGATCTGCTCGAGTGAAGAACTTCTCG 20767
QY 201 GlnSerLeuArgAlaAlaGlyPheGluIleLeuAspTrpGluAspValSerSerArgThr 220
DB 20768 GAATGCTGCTGCGACGCGGGTTCAGATCTTCGATTCGGAGAGGACGTGTCTCGAGGACC 20827
QY 221 ArgTyrPheMetProGlnPheAlaGluGluLeuAlaAlaHisGlnHisGlyIleAlaAsp 240
DB 20828 CGGTACTTCATGCCACAGTTCGCCGAAGAGCTCGCTGGCACACACGCGGATCCGGAC 20887
QY 241 ArgTyrArgIleProAlaValAlaGlyTyrPheAlaAlaAlaValCysAspTyrGluLysTyrAla 260
DB 20888 AGGTACGGGCGCGGTGTGCGCGGCTGGGCGCGGCTTCGATTATGAGAAATATATGCC 20947
```

```
QY 261 HisAspMetGlyTyrAlaIleLeuThrAlaArgIleYspProValGly 275
DB 20948 CACGACATGGGCTATGCGATTTCTACGCGCGGAGAACCGGTGCGG 20992

RESULT 2
US-09-370-700-1
/ Sequence 1, Application US/09370700
/ Patent No. 6274350
/ GENERAL INFORMATION:
/ APPLICANT: Balcz, Richard H
/ APPLICANT: Broughton, Mary C
/ APPLICANT: Crawford, Kathryn P
/ APPLICANT: Madhuri, Krishnamurthy
/ APPLICANT: Treadway, Patti J
/ APPLICANT: Turner, Jan R
/ APPLICANT: Waldron, Clive
/ TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
/ FILE REFERENCE: 50489 DIV1
/ CURRENT APPLICATION NUMBER: US/09/370,700
/ CURRENT FILING DATE: 1999-08-09
/ EARLIER APPLICATION NUMBER: US 09/36987
/ EARLIER FILING DATE: 1998-03-09
/ NUMBER OF SEQ ID NOS: 39
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 1
/ LENGTH: 80161
/ TYPE: DNA
/ ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Alignment Scores:
Pred. No.: 5,22e-273 Length: 80161
Score: 275.00 Matches: 275
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-069-353A-8 (1-275) x US-09-370-700-1 (1-80161)

QY 1 ValLeuProGlyValAlaProThrSerGlnGlnValGlyGlnMetTyrAspLeuValThr 20
DB 20168 GTGTTGCCAGGTGGCGCACCAATCGCAGGTTGGGCAATGATGATGACCTGTGTCACG 20227
QY 21 ProLeuLeuAsnSerValAlaGlyGlyProCysAlaIleHisGlyTyrTrpGluAsn 40
DB 20228 CCGTTCCTGAATCTCGGTCCGGCCGGCCCTTCGCGCATTCACACGGCTACTGGAGAAC 20287
QY 41 AspGlyArgAlaSerTrpGlnGlnAlaAlaAspArgLeuThrAspLeuValAlaGluArg 60
DB 20288 GACGGCGCGGCTTCTCGTGGCAGAGCGCGCCGACCGGCTCACCGACTTGTCCGCAACGG 20347
QY 61 ThrValLeuAspGlyValAlaArgLeuLeuAspValGlyCysGlyThrGlyGlnProAla 80
DB 20348 ACCGTGCTCGATGGCGGCTTGCACCTGCTGATGTGGGGTGGGTAACCGAACACGCGG 20407
QY 81 LeuArgValAlaArgAspAsnAlaIleGlnIleThrGlyIleThrValSerGlnValGln 100
DB 20408 CTGGCGCTGCGCGCGACCAACCGCATCCGATCACCGGCTACCGGAGAGTGCNA 20467
QY 101 ValAlaIleAlaAlaAspCysAlaArgGluArgGlyLeuSerHisArgValAspPheSer 120
DB 20468 GTGGCCATGCGCGCATGATGGCGACGCGGACGCGGATPAAGCCACCGGGTGGACTTCTCG 20527
QY 121 CysValAspAlaMetSerLeuProTyrProAspAsnAlaPheAspAlaAlaTrpAlaMet 140
DB 20528 TGCCTCGATGCAATGCTCTGCGCGTACCGGACCAATGCTTTCGACGCGCTGGGCGCATG 20587
QY 141 GlnSerLeuLeuGlnMetSerGluProAspArgAlaIleArgGluIleLeuArgValLeu 160
DB 20588 CAGTGGCTGTGGAGATGTCGAACCGGACCGGCTGCATTCGGGAAATCTTTCGATGACTC 20647
```







COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,986A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benson, Bob  
REGISTRATION NUMBER: 30,446  
REFERENCE/DOCKET NUMBER: PB248PP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 2979:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 400 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-781-986A-2979

Alignment Scores:  
Pred. No.: 3.89 Length: 400  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.27% Indels: 0  
DB: 4 Gaps: 0

US-10-069-353a-8 (1-275) x US-08-781-986A-2979 (1-400)

QY 158 ArgValIleuLysProGlyGlyIleLeu 166  
Db 151 CGAGCTTAAAGCCTGGGGGTATCTTG 125

RESULT 8  
US-08-956-171E-444  
Sequence 444, Application US/08956171E  
Patent No. 6593114  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
Gil H. Choi  
Patrick S. Dillon  
Craig A. Rosen  
Steven C. Barash  
Michael R. Fannon  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5256  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/956,171E  
FILING DATE: 20-Oct-1997  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/009,861  
FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986  
FILING DATE: January 3, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mark J. Hyman  
REGISTRATION NUMBER: 46,789  
REFERENCE/DOCKET NUMBER: PB248P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (240) 314-1224  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 444:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11466 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-956-171E-444

Alignment Scores:  
Pred. No.: 94.9 Length: 11466  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.27% Indels: 0  
DB: 4 Gaps: 0

US-10-069-353a-8 (1-275) x US-08-956-171E-444 (1-11466)

QY 158 ArgValIleuLysProGlyGlyIleLeu 166  
Db 10859 CGAGCTTAAAGCCTGGGGGTATCTTG 10885

RESULT 9  
US-08-781-986A-444  
Sequence 444, Application US/08781986A  
Patent No. 6737248  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,986A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benson, Bob  
REGISTRATION NUMBER: 30,446  
REFERENCE/DOCKET NUMBER: PB248PP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 444:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11466 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-781-986A-444

## Alignment Scores:

Pred. No.: 94.9 Length: 11466  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.27% Indels: 0  
 DB: 4 Gaps: 0

US-10-069-353A-8 (1-275) x US-08-781-986A-444 (1-11466)

QY 158 ArgValLeuLysProGlyGlyLeu 166

DB 10859 CGAGCTTAAGCTGGGGGATCTTG 10885

## RESULT 10

US-09-270-767-28714  
 ; Sequence 28714, Application US/09270767  
 ; Patent No. 6703491  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Homburger et al.  
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
 ; FILE REFERENCE: File Reference: 7326-094  
 ; CURRENT APPLICATION NUMBER: US/09/270,767  
 ; PRIORITY FILING DATE: 1999-03-17  
 ; NUMBER OF SEQ ID NOS: 62517  
 ; SOFTWARE: Patent Ver. 2.0  
 ; SEQ ID NO 28714  
 ; LENGTH: 310  
 ; TYPE: DNA  
 ; ORGANISM: Drosophila melanogaster  
 US-09-270-767-28714

## Alignment Scores:

Pred. No.: 33 Length: 310  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.91% Indels: 0  
 DB: 4 Gaps: 0

US-10-069-353A-8 (1-275) x US-09-270-767-28714 (1-310)

QY 70 LeuAspValGlyCysGlyThrGly 77

DB 48 TTGGATGTGGATGCCGCAAGGGA 71

## RESULT 11

US-09-252-991A-334/C  
 ; Sequence 334, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIORITY FILING DATE: 1999-02-18  
 ; NUMBER OF SEQ ID NOS: 60/074,788  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIORITY FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 334  
 ; LENGTH: 414  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-334

## Alignment Scores:

Pred. No.: 43.5 Length: 414  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.91% Indels: 0  
 DB: 4 Gaps: 0

US-10-069-353A-8 (1-275) x US-09-252-991A-334 (1-414)

QY 157 LeuArgValLeuLysProGlyGly 164

DB 44 CTGAGGTCCTCAAGCCCGCGGC 21

## RESULT 12

US-09-248-796A-3885  
 ; Sequence 3885, Application US/09248796A  
 ; Patent No. 6747137  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Keith Weinstein et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
 ; FILE REFERENCE: 107196.132  
 ; CURRENT APPLICATION NUMBER: US/09/248,796A  
 ; PRIORITY FILING DATE: 1999-02-12  
 ; NUMBER OF SEQ ID NOS: 28208  
 ; SEQ ID NO 3885  
 ; LENGTH: 456  
 ; TYPE: DNA  
 ; ORGANISM: Candida albicans  
 US-09-248-796A-3885

## Alignment Scores:

Pred. No.: 47.7 Length: 456  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.91% Indels: 0  
 DB: 4 Gaps: 0

US-10-069-353A-8 (1-275) x US-09-248-796A-3885 (1-456)

QY 70 LeuAspValGlyCysGlyThrGly 77

DB 199 CTTGATGTTGGATGTGATCTGT 222

## RESULT 13

US-09-621-976-18024/C  
 ; Sequence 18024, Application US/09621976  
 ; Patent No. 6639063  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Milne Edwards, J.B.  
 ; APPLICANT: Jobert, S.  
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
 ; FILE REFERENCE: GENSET.054PR2  
 ; CURRENT APPLICATION NUMBER: US/09/621,976  
 ; PRIORITY FILING DATE: 2000-07-21  
 ; NUMBER OF SEQ ID NOS: 19335  
 ; SOFTWARE: Patent.pm  
 ; SEQ ID NO 18024  
 ; LENGTH: 505  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-621-976-18024

## Alignment Scores:

Pred. No.: 52.6 Length: 505  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.91% Indels: 0  
 DB: 4 Gaps: 0

US-10-069-353A-8 (1-275) x US-09-621-976-18024 (1-505)

QY 244 ProAlaValAlaGlyTrpAlaAla 251  
|||||  
DB 296 CCGGACGTTGGCTGGGCGGACGACA 273

RESULT 14  
US-09-902-540-7484

; Sequence 7484, Application US/09902540

; Patent No. 6833447

; GENERAL INFORMATION:

; APPLICANT: Goldman, Barry S.

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Wiegand, Roger C.

; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

; FILE REFERENCE: 38-10(15849)B

; CURRENT APPLICATION NUMBER: US/09/902,540

; PRIOR FILING DATE: 2001-07-10

; PRIOR APPLICATION NUMBER: 60/217,883

; NUMBER OF SEQ ID NOS: 16825

; SEQ ID NO 7484

; LENGTH: 507

; TYPE: DNA

; ORGANISM: Myxococcus xanthus

US-09-902-540-7484

Alignment Scores:

Pred. No.:

52.8

Length:

507

Score:

8.00

Matches:

8

Percent Similarity:

100.00%

Conservative:

0

Best Local Similarity:

100.00%

Mismatches:

0

Query Match:

2.91%

Indels:

0

DB: 4 Gaps: 0

US-10-069-353A-8 (1-275) x US-09-902-540-7484 (1-507)

QY 70 LeuAspValAlaGlyCysGlyThrGly 77  
|||||

DB 28 TTGGACGTGGCTGGGCGGACGCGG 51

RESULT 15  
US-09-902-540-3761

; Sequence 3761, Application US/09902540

; Patent No. 6833447

; GENERAL INFORMATION:

; APPLICANT: Goldman, Barry S.

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Wiegand, Roger C.

; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

; FILE REFERENCE: 38-10(15849)B

; CURRENT APPLICATION NUMBER: US/09/902,540

; PRIOR FILING DATE: 2001-07-10

; PRIOR APPLICATION NUMBER: 60/217,883

; NUMBER OF SEQ ID NOS: 16825

; SEQ ID NO 3761

; LENGTH: 522

; TYPE: DNA

; ORGANISM: Myxococcus xanthus

US-09-902-540-3761

Alignment Scores:

Pred. No.:

54.2

Length:

522

Score:

8.00

Matches:

8

Percent Similarity:

100.00%

Conservative:

0

Best Local Similarity:

100.00%

Mismatches:

0

Query Match:

2.91%

Indels:

0

DB: 4 Gaps: 0

US-10-069-353A-8 (1-275) x US-09-902-540-3761 (1-522)

QY 70 LeuAspValAlaGlyCysGlyThrGly 77  
|||||  
DB 58 CTGGACGTTGGCTGGGCGGACCTGG 81

Search completed: August 30, 2005, 01:08:17  
Job time : 288 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 29, 2005, 22:45:45 ; Search time 699 Seconds

(without alignments)  
2574.177 Million cell updates/sec

Title: US-10-069-353a-8

Perfect score: 275

Sequence: 1 VLPGGAPTSQQVGMIDLVY.....YEKYADMGYAILTARKPVG 275

Scoring table:

OLIGO	Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 7331713 segs, 3271544945 residues

Word size: 1

Total number of hits satisfying chosen parameters: 14651541

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

```
-MODE=frame+ p2n.model -DEV=x1h
-O/cgn2_1/USPTO.spool/US10069353.rnuc 26082005.172153.15107/app.query.fasta_1.455
-DB=Published Applications NA -QFMT=fastap -SUPFT=rmdb -MINMATCH=0.1
-LOOPCU=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cgi -LIST=45 -DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10069353@cgn 1 1 480 @rnuc 26082005.172153.15107
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -MAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YCAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7
```

Database : Published Applications NA:\*

```
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PTC_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/PTCUS_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	275	100.0	80161	17	US-10-329-148A-1	Sequence 1, Appli
2	9	3.3	400	8	US-08-781-986A-2979	Sequence 2979, Ap
3	9	3.3	400	18	US-10-329-624-2979	Sequence 2979, Ap
4	9	3.3	11466	8	US-08-781-986A-444	Sequence 444, App
5	9	3.3	11466	18	US-10-329-624-444	Sequence 444, App
6	9	3.3	47745	13	US-10-087-132-469	Sequence 469, App
7	9	3.3	64492	18	US-10-378-083-1	Sequence 1, Appli
8	9	2.9	262	20	US-10-425-115-21054	Sequence 21054, A
9	8	2.9	270	9	US-09-954-456-1002	Sequence 1002, Ap
10	8	2.9	270	10	US-09-960-706-835	Sequence 835, App
11	8	2.9	270	10	US-09-873-319-539	Sequence 539, App
12	8	2.9	270	17	US-10-170-385-168	Sequence 168, App
13	8	2.9	270	21	US-10-843-641A-4029	Sequence 4029, App
14	8	2.9	318	20	US-10-425-115-55334	Sequence 55334, A
15	8	2.9	333	15	US-10-156-761-81	Sequence 81, Appl
16	8	2.9	340	20	US-10-425-115-154357	Sequence 154357,
17	8	2.9	344	20	US-10-425-115-138703	Sequence 138703,
18	8	2.9	349	20	US-10-357-930-7576	Sequence 9363, Ap
19	8	2.9	381	20	US-10-425-115-9363	Sequence 9363, Ap
20	8	2.9	405	20	US-10-357-930-37536	Sequence 37536, A
21	8	2.9	441	21	US-10-472-928-2613	Sequence 2613, Ap
22	8	2.9	459	19	US-10-474-776-518	Sequence 518, App
23	8	2.9	483	10	US-09-918-995-617	Sequence 617, App
24	8	2.9	484	10	US-09-918-995-29551	Sequence 29551, A
25	8	2.9	513	18	US-10-424-599-130420	Sequence 130420,
26	8	2.9	514	9	US-09-920-300A-1637	Sequence 1637, Ap
27	8	2.9	514	13	US-10-033-528-1637	Sequence 1637, Ap
28	8	2.9	514	16	US-10-099-926-1637	Sequence 1637, Ap
29	8	2.9	514	22	US-10-961-527-1637	Sequence 1637, Ap
30	8	2.9	528	17	US-10-259-194A-625	Sequence 625, App
31	8	2.9	561	13	US-10-027-632-32254	Sequence 32254, App
32	8	2.9	561	17	US-10-027-632-32254	Sequence 32254, App
33	8	2.9	564	19	US-10-767-701-29067	Sequence 29067, A
34	8	2.9	576	18	US-10-424-599-138183	Sequence 138183,
35	8	2.9	579	9	US-09-810-936-147	Sequence 147, App
36	8	2.9	579	9	US-09-429-755-147	Sequence 147, App
37	8	2.9	579	9	US-09-924-400-147	Sequence 147, App
38	8	2.9	579	15	US-10-212-679-147	Sequence 147, App
39	8	2.9	579	18	US-10-079-137B-147	Sequence 147, App
40	8	2.9	599	22	US-10-972-079-31973	Sequence 31973, A
41	8	2.9	600	22	US-10-972-079-3372	Sequence 3272, Ap
42	8	2.9	600	22	US-10-972-079-3373	Sequence 3273, Ap
43	8	2.9	600	22	US-10-972-079-3274	Sequence 3274, Ap
44	8	2.9	600	22	US-10-972-079-3275	Sequence 3275, Ap
45	8	2.9	600	22	US-10-972-079-31974	Sequence 31974, A

## ALIGNMENTS

```
RESULT 1
US-10-329-148A-1
; Sequence 1, Application US/10329148A
; Publication No. US20040023343A1
GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/10/329,148A
; CURRENT FILING DATE: 2002-12-23
```

[illegible]

```

Oy      261 HispmetGlyTyrAlaIleLeuThrAlaArgLysProValGly 275
      |||
Db      20948 CACGACATGGGCTATGCGATTCTGACGGCGCGAAGCCGCTCGGC 20992

RESULT 2
US-08-781-986A-2979/c
; Sequence 2979, Application US/08781986A
; Publication No. US20030054436A1
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SRO ID NO: 2979:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-2979

Alignment Scores:
Pred. No.: 19.5 Length: 400
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.27% Indels: 0
DB: 8 Gaps: 0

US-10-069-353A-8 (1-275) x US-08-781-986A-2979 (1-400)
Oy      158 ArgValleuLysProGlyGlyIleLeu 166
      |||
Db      151 CGAGCTTAAAGCCTGAGGGGTATCTTG 125

RESULT 3
US-10-329-624-2979/c
; Sequence 2979, Application US/10329624
; Publication No. US20040043037A1
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
GIL H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

```



NUMBER OF SEQUENCES: 5256  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/329,624  
FILING DATE: 27-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/956,171  
FILING DATE: October 20, 1997  
APPLICATION NUMBER: 60/009,861  
FILING DATE: January 5, 1996  
APPLICATION NUMBER: 08/781,986  
FILING DATE: January 3, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mark J. Hyman  
REGISTRATION NUMBER: 46,789  
REFERENCE/DOCKET NUMBER: PB248P1D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (240) 314-1224  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 2979:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 400 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 2979:  
US-10-329-624-2979  
Alignment Scores:  
Pred. No.: 19.5 Length: 400  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.27% Indels: 0  
DB: 18 Gaps: 0  
US-10-069-353a-8 (1-275) x US-10-329-624-2979 (1-400)  
QY 158 ArgValIeuIysProGlyGIyIleLeu 166  
Db 151 CGAGTCTTAAAGCTCGGGGTATCTTG 125  
RESULT 4  
US-08-781-986A-444  
Sequence 444, Application US/08781986A  
Publication No. US20030054436A1  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5256  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,986A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benson, Bob  
REGISTRATION NUMBER: 30,446  
REFERENCE/DOCKET NUMBER: PB248PP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 444:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11466 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-781-986A-444  
Alignment Scores:  
Pred. No.: 339 Length: 11466  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.27% Indels: 0  
DB: 8 Gaps: 0  
US-10-069-353a-8 (1-275) x US-08-781-986A-444 (1-11466)  
QY 158 ArgValIeuIysProGlyGIyIleLeu 166  
Db 10859 CGAGTCTTAAAGCTCGGGGTATCTTG 10885  
RESULT 5  
US-10-329-624-444  
Sequence 444, Application US/10329624  
Publication No. US20040043037A1  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
Gil H. Choi  
Patrick S. Dillon  
Craig A. Rosen  
Steven C. Barash  
Michael R. Fannon  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5256  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/329,624  
FILING DATE: 27-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/956,171  
FILING DATE: October 20, 1997  
APPLICATION NUMBER: 60/009,861  
FILING DATE: January 5, 1996  
APPLICATION NUMBER: 08/781,986  
FILING DATE: January 3, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mark J. Hyman

```
/
/ REGISTRATION NUMBER: 46,789
/ REFERENCE/DOCKET NUMBER: PB248PID1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (240) 314-1224
/ TELEFAX: (301) 309-8439
/ INFORMATION FOR SEQ ID NO: 444:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 11466 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ US-10-329-624-444
/ SEQUENCE DESCRIPTION: SEQ ID NO: 444:

Alignment Scores:
Pred. No.: 339 Length: 11466
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.27% Indels: 0
DB: 18 Gaps: 0

US-10-069-353A-8 (1-275) x US-10-329-624-444 (1-11466)

QY 158 ArgValLeuValProGlyGlyIleLeu 166
DB 10859 CGAGCTTTAAAGCTGGGGGTATCTTG 10885

RESULT 6
US-10-087-192-469/c
/ Sequence 469, Application US/10087192
/ Publication No. US20020182586A1
/ GENERAL INFORMATION:
/ APPLICANT: Morris, David W.
/ APPLICANT: Engelhard, Eric K.
/ TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
/ FILE REFERENCE: 529452000122
/ CURRENT APPLICATION NUMBER: US/10/087,192
/ CURRENT FILING DATE: 2002-03-01
/ PRIOR APPLICATION NUMBER: US 09/747,377
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: US 09/798,586
/ PRIOR FILING DATE: 2001-03-02
/ NUMBER OF SEQ ID NOS: 2059
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 469
/ LENGTH: 47745
/ TYPE: DNA
/ ORGANISM: Mus musculus
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(47745)
/ OTHER INFORMATION: n = A,T,C or G
US-10-087-192-469

Alignment Scores:
Pred. No.: 1.14e+03 Length: 47745
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.27% Indels: 0
DB: 13 Gaps: 0

US-10-069-353A-8 (1-275) x US-10-087-192-469 (1-47745)

QY 56 LeuValAlaGluArgThrValLeuAsp 64
DB 35691 TTAGTGGCAGAGAGACTGTACTTAC 35665

RESULT 7
US-10-378-083-1/c
/ Sequence 1, Application US/10378083
```

```
/ Publication No. US20040053274A1
/ GENERAL INFORMATION:
/ APPLICANT: President of Tokyo Institute of Technology
/ TITLE OF INVENTION: A gene cluster of vicenistatin biosynthase, a vicenistamine
/ TITLE OF INVENTION: glycosyltransferase polypeptide, and a gene encoding the
/ FILE REFERENCE: polypeptide
/ CURRENT APPLICATION NUMBER: US/10/378,083
/ CURRENT FILING DATE: 2003-03-04
/ NUMBER OF SEQ ID NOS: 26
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 1
/ LENGTH: 64492
/ TYPE: DNA
/ ORGANISM: Streptomyces halstedii sp.HC-34
US-10-378-083-1

Alignment Scores:
Pred. No.: 1.47e+03 Length: 64492
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.27% Indels: 0
DB: 18 Gaps: 0

US-10-069-353A-8 (1-275) x US-10-378-083-1 (1-64492)

QY 69 LeuLeuAspValGlyCysGlyThrGly 77
DB 33852 CTGCTCGATGTGGCTGTGGCACCGGC 33826

RESULT 8
US-10-425-115-21054/c
/ Sequence 21054, Application US/10425115
/ Publication No. US20040214272A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
/ FILE REFERENCE: 38-21(53222)B
/ CURRENT APPLICATION NUMBER: US/10/425,115
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 369326
/ SEQ ID NO 21054
/ LENGTH: 262
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(262)
/ OTHER INFORMATION: unsure at all n locations
/ OTHER INFORMATION: Clone ID: WRT4577_119199C.1
US-10-425-115-21054

Alignment Scores:
Pred. No.: 145 Length: 262
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.91% Indels: 0
DB: 20 Gaps: 0

US-10-069-353A-8 (1-275) x US-10-425-115-21054 (1-262)

QY 63 LeuAspGlyValArgLeuLeu 70
DB 47 CTGACGGGGCGTTCCTCTC 24

RESULT 9
```

```
US-09-954-456-1002/c
; Sequence 1002, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1002
; LENGTH: 270
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: n=a,t,g or c
US-09-954-456-1002

Alignment Scores:
Pred. No.: 149 Length: 270
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.91% Indels: 0
DB: 9 Gaps: 0

US-10-069-353a-8 (1-275) x US-09-954-456-1002 (1-270)

QY 17 AspleuValThProleuLeuAn 24
Db 92 GATCTGTCACCTCTTGTCTCAAC 69

RESULT 10
US-09-960-706-835/c
; Sequence 835, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplas
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 835

US-10-069-353a-8 (1-275) x US-09-960-706-835 (1-270)

QY 17 AspleuValThProleuLeuAn 24
Db 92 GATCTGTCACCTCTTGTCTCAAC 69

RESULT 11
US-09-873-319-539/c
; Sequence 539, Application US/09873319A
; Publication No. US20030134324A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; APPLICANT: Kulkarni, Prakash
; APPLICANT: Getzenberg, Robert H.
; APPLICANT: Waga, Iwao
; APPLICANT: Yamamoto, Jun
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic
; FILE REFERENCE: 44921-5029-US
; CURRENT APPLICATION NUMBER: US/09/873,319A
; CURRENT FILING DATE: 2001-06-05
; EARLIER APPLICATION NUMBER: US 60/223,323
; EARLIER FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 755
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 539
; LENGTH: 270
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 R44397
; NAME/KEY: unsure
; LOCATION: (1)..(270)
; OTHER INFORMATION: n = a or c or g or t
US-09-873-319-539

Alignment Scores:
Pred. No.: 149 Length: 270
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.91% Indels: 0
DB: 10 Gaps: 0

US-10-069-353a-8 (1-275) x US-09-873-319-539 (1-270)

QY 17 AspleuValThProleuLeuAn 24
Db 92 GATCTGTCACCTCTTGTCTCAAC 69

RESULT 12
US-10-170-385-168/c
; Sequence 168, Application US/10170385
; Publication No. US2003020372A1
```

```
/ GENERAL INFORMATION:
/ APPLICANT: Ward, Neil Raymond
/ APPLICANT: Mundy, Christopher Robert
/ APPLICANT: Kan, On
/ APPLICANT: Harris, Robert Alan
/ APPLICANT: White, Jonathan
/ APPLICANT: Binley, Katie Mary
/ APPLICANT: Rayner, William Nigel
/ APPLICANT: Naylor, Stuart
/ APPLICANT: Kingsman, Susan Mary
/ APPLICANT: Krige, David
/ TITLE OF INVENTION: ANALYSIS METHOD
/ FILE REFERENCE: 532682000100
/ CURRENT APPLICATION NUMBER: US/10/170,385
/ CURRENT FILING DATE: 2002-06-12
/ PRIOR APPLICATION NUMBER: PCT/GB02/01662
/ PRIOR FILING DATE: 2002-04-08
/ PRIOR APPLICATION NUMBER: PCT/GB01/05458
/ PRIOR FILING DATE: 2001-12-10
/ NUMBER OF SEQ ID NOS: 549
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 168
/ LENGTH: 270
/ TYPE: DNA
/ ORGANISM: Homo Sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 144..233, 248
/ OTHER INFORMATION: n = A,T,C or G
US-10-170-385-168
```

```
Alignment Scores:
Pred. No.: 149 Length: 270
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.91% Indels: 0
DB: 17 Gaps: 0
```

US-10-069-353A-8 (1-275) x US-10-170-385-168 (1-270)

```
QY 17 AspleuValThProLeuLeuAsn 24
|||
Db 92 GATCTCGTCACTCCTTGTCTCAAC 69
```

```
RESULT 13
US-10-843-641A-4029/C
/ Sequence 4029, Application US/10843641A
/ Publication No. US20050064454A1
/ GENERAL INFORMATION:
/ APPLICANT: Avalon Pharmaceuticals, Inc.
/ TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
/ FILE REFERENCE: 689290-189
/ CURRENT APPLICATION NUMBER: US/10/843,641A
/ CURRENT FILING DATE: 2004-05-12
/ PRIOR APPLICATION NUMBER: US/09/873,367
/ PRIOR FILING DATE: 2001-06-05
/ PRIOR APPLICATION NUMBER: US/09/954,531
/ PRIOR FILING DATE: 2001-09-18
/ PRIOR APPLICATION NUMBER: US/09/954,456
/ PRIOR FILING DATE: 2001-09-25
/ PRIOR APPLICATION NUMBER: US/09/962,436
/ PRIOR FILING DATE: 2001-09-25
/ PRIOR APPLICATION NUMBER: US/09/962,832
/ PRIOR FILING DATE: 2001-09-25
/ PRIOR APPLICATION NUMBER: US/09/964,824
/ PRIOR FILING DATE: 2001-09-27
/ PRIOR APPLICATION NUMBER: US/09/967,768
/ PRIOR FILING DATE: 2001-09-28
/ PRIOR APPLICATION NUMBER: US/09/968,007
/ PRIOR FILING DATE: 2001-10-02
/ PRIOR APPLICATION NUMBER: US/09/969,347
```

```
/ PRIOR FILING DATE: 2001-10-02
/ PRIOR APPLICATION NUMBER: US/09/969,708
/ PRIOR FILING DATE: 2001-10-03
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 8447
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO: 4029
/ LENGTH: 270
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)..(270)
/ OTHER INFORMATION: n=a,t,g or c
US-10-843-641A-4029
```

```
Alignment Scores:
Pred. No.: 149 Length: 270
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.91% Indels: 0
DB: 21 Gaps: 0
```

US-10-069-353A-8 (1-275) x US-10-843-641A-4029 (1-270)

```
QY 17 AspleuValThProLeuLeuAsn 24
|||
Db 92 GATCTCGTCACTCCTTGTCTCAAC 69
```

```
RESULT 14
US-10-425-115-55334
/ Sequence 55334, Application US/10425115
/ Publication No. US20040214272A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(53222)B
/ CURRENT APPLICATION NUMBER: US/10/425,115
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 369326
/ SEQ ID NO: 55334
/ LENGTH: 318
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ OTHER INFORMATION: Clone ID: MRT4577_150460C.1
US-10-425-115-55334
```

```
Alignment Scores:
Pred. No.: 171 Length: 318
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.91% Indels: 0
DB: 20 Gaps: 0
```

US-10-069-353A-8 (1-275) x US-10-425-115-55334 (1-318)

```
QY 162 ProgiygiyleLeuGlyValThr 169
|||
Db 208 CCCGCGGTATATCGGGTGTAAAC 231
```

```
RESULT 15
US-10-156-761-81
/ Sequence 81, Application US/10156761
/ Publication No. US20030119018A1
/ GENERAL INFORMATION:
/ APPLICANT: OMURA, SATOSHI
```

APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, TUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO: 81  
LENGTH: 333  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(333)  
US-10-156-761-81

Alignment Scores:  
Pred. No.: 178 Length: 333  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.91% Indels: 0  
DB: 15 Gaps: 0

US-10-069-353a-8 (1-275) x US-10-156-761-81 (1-333)

Qy 70 LeuAspValGlyCysGlyThrGly 77  
Db 295 CTGGACCTGGGCTGCCGACGGGC 318

Search completed: August 30, 2005, 01:19:56  
Job time : 748 secs

This Page Blank (uspto)